

# **Protein Allostery: Computational Characterization of Diguanylate Cyclase PleD**

**Inauguraldissertation**

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch–Naturwissenschaftlichen Fakultät

der Universität Basel

von

Franziska F.-F. Schmid

aus Wittnau, AG

Basel, 2009

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät

auf Antrag von:

Prof. Dr. Markus Meuwly

Prof. Dr. Lennart Nilsson

Basel, den 11. November 2008

Prof. Dr. E. Parlow (Dekan)

# Acknowledgment

I would like to thank my supervisor Prof. Markus Meuwly for providing me with this very interesting project. I appreciate his guidance and the helpful discussions throughout my PhD study. I would also like to thank Profs. Urs Jenal and Tilman Schirmer who initiated the collaboration and provided the protein structure and biochemical data. It has been a wonderful experience to see the joint effort of different scientific fields to shed light on the c-di-GMP puzzle.

I would also like to thank Prof. Lennart Nilsson who kindly accepted to act as external examiner.

It has been a pleasure to work with the members of the Meuwly group. In particular, I am indebted to Dr. David Nutt for introducing me to the world of computational science and I enjoyed sharing the office with Manuela Koch. English proof-reading of the thesis by Dr. Michael Devereux is acknowledged. Further, I would like to thank Dr. Vincent Zoete, Dr. Polina Banushkina, Dr. Sven Lammers, Dr. Jonas Danielsson, Dr. Antony Fouqueau, Dr. Ivan Tubert-Brohman, Dr. Sabyashachi Mishra, Hólmfríður Þorsteinsdóttir, Nuria Plattner, Stephan Lutz, Jing Huang, Tobias Schmidt and Maurus Schmid.

I would like to thank Matthias and Beat Christen for their stimulating discussions and their efforts in teaching molecular biology.

Last but not least I would like to thank my family for their unconditional support.



# Abstract

In biology, accurate cellular regulation in response to environmental signals is crucial for the fitness of organisms. On the molecular level the modulation of protein activity is often achieved by the binding of a signaling molecule or by covalent modifications such as phosphorylation by kinases. Protein allostery, that is signal propagation from a distant allosteric site to functional sites to regulate the output, has long been recognized but the views and perspectives have been strongly influenced by different scientific fields and the continuous development of new methods. In particular, computational approaches are suited to bridge the gap between structure and dynamics and provide insight at an atomic level.

After reviewing experimental and theoretical methods to study allostery, results from computational methods applied to the diguanylate cyclase PleD are presented. First, structural and dynamical aspects of the communication between the allosteric inhibition site and the active site are highlighted by energy calculations and molecular dynamics simulations. Ligand binding may trigger a balance-like movement of the conserved strand  $\beta 2$  that potentially displaces residues required for catalysis. In addition, dynamical coupling between the functional sites, i.e. simultaneous quenching of motion upon ligand binding, is found from normal mode analysis. Furthermore, two possible communication pathways connecting the inhibition with the active site are proposed. Second, processes involved in PleD dimerization were elucidated. In dynamics simulations the spontaneous active-to-inactive transition is observed and implies changes in the D1/D2 interface together with a slight decrease in the dimerization contact area. In the proposed model the  $\beta 4$ - $\alpha 4$  loop repositioning is followed by adjustments in the  $\alpha 4$ - $\beta 5$ - $\alpha 5$  face that are amplified by the extended helix  $\alpha 5$  by a leverage effect.



# Contents

<b>Acknowledgment</b>	<b>i</b>
<b>Abstract</b>	<b>iii</b>
<b>Abbreviations</b>	<b>vii</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Protein Allostery</b>	<b>3</b>
<b>3 Methods Survey</b>	<b>9</b>
3.1 Elucidating biological function and pathways . . . . .	9
3.2 Structure determination . . . . .	13
3.2.1 X-ray crystallography . . . . .	15
3.2.2 Cryo-electron microscopy . . . . .	16
3.3 Protein dynamics . . . . .	17
3.3.1 Nuclear magnetic resonance (NMR) spectroscopy . . . . .	18
3.3.2 Fluorescence-resonance energy transfer (FRET) . . . . .	21
3.4 Thermodynamics . . . . .	24
3.4.1 Isothermal titration calorimetry (ITC) . . . . .	24
3.5 Computational contributions to mechanistic understanding . . . . .	24
3.5.1 Root-mean square deviation (RMSD) . . . . .	25
3.5.2 Root-mean square fluctuation (RMSF) . . . . .	26
3.5.3 Dynamic cross-correlated motion . . . . .	27
3.5.4 Normal mode analysis (NMA) . . . . .	28
3.5.5 Biased molecular dynamics simulations . . . . .	31
3.5.6 Domain identification . . . . .	33
3.5.7 Calculation of energy contributions and hot-spot residue identification . . . . .	35

3.5.8	Bioinformatics: sequence-structure relationship . . . . .	35
<b>4</b>	<b>Bacterial Second Messenger c-di-GMP and PleD as Model System</b>	<b>41</b>
4.1	Ubiquitous bacterial second messenger c-di-GMP . . . . .	41
4.2	PleD activity under allosteric control . . . . .	43
<b>5</b>	<b>Results</b>	<b>49</b>
5.1	Allosteric control of cyclic di-GMP signaling . . . . .	49
5.2	All-atom simulations of structures and energetics of c-di-GMP bound and free PleD . . . . .	68
5.3	PleD dimerization process . . . . .	90
<b>6</b>	<b>Discussion</b>	<b>105</b>
<b>7</b>	<b>Future Directions</b>	<b>109</b>
	<b>Appendix</b>	<b>129</b>



# Abbreviations

A-site	active site
ATD	anisotropic thermal diffusion
c-di-GMP	bis-(3'-5')-cyclic dimeric guanosine monophosphate
CheY	chemotaxis protein Y
DCCM	dynamical cross-correlation map
DGC	diguanylate cyclase
DgcA	diguanylate cyclase A (CC3285)
EAL domain	Glu-Ala-Leu domain harboring c-di-GMP specific PDE activity
EM	electron microscopy
ENM	elastic network model
FRET	fluorescence resonance energy transfer
GGDEF domain	Gly-Gly-Asp-Glu-Phe domain harboring DGC activity
GNM	Gaussian network model
GTP	guanosine triphosphate
I-site	inhibition site
ITC	isothermal titration calorimetry
KNF	Koshland, Nemethy and Filmer
MD	molecular dynamics
MWC	Monod, Wyman and Changeux
NM(A)	normal mode (analysis)
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
PDB	protein data bank
PDE	phosphodiesterase
PleD	DGC protein involved in lifecycle of <i>C. crescentus</i>
RDC	residual dipolar coupling
REC	receiver domain
RMSD	root mean-square deviation
RMSF	root mean-square fluctuation

RR	response regulator
SMD	steered molecular dynamics simulations
TMD	targeted molecular dynamics simulations
wt	wildtype

# 1 Introduction

In biology, cellular regulation and control is crucial for organisms to be able to respond to drastic changes in their extracellular environments, e.g. nutrient availability, and requires adaptive responses to altered conditions. On the molecular level, this modulation of protein activity is often achieved by binding of an effector molecule, i.e. signaling molecules which enhance or inhibit the function of a certain protein. Local perturbation by the ligand binding to the protein at an allosteric site influences the activity of another, often spatially distant site. Protein allostery is the process of propagating a structural and/or dynamical change from a specific site throughout the protein and thereby regulating its output, e.g. catalytic activity or further protein-molecule interactions. Detailed understanding of protein function, including the processes of information transfer in protein allostery, is important in order to manipulate signaling cascades and provide a basis for finding new potential targets in antimicrobial drug development.

Protein allostery has long been recognized and has been studied extensively in the last decades with a variety of experimental and theoretical techniques developed in the fields of physics, biology, chemistry and computer sciences. In the course of time the views on protein allostery have changed and were substantially extended due to influences from various scientific fields. The old-established notion of a change in protein shape, based on the induced-fit concept, and mechanistic structure rearrangement is supported by the comparison of available structures of liganded and unliganded proteins. Yet, increasing data from nuclear magnetic resonance (NMR) studies and time-resolved methods emphasize the importance of protein motion and dynamics. In chapter 2 opinions on protein allostery are summarized and a survey of methods, describing the specific contributions and advantages of numerous experimental and theoretical tools, is given in chapter 3. The different approaches to study signal transduction and allosteric mechanisms are illustrated with examples. The main focus is on the impact of computational investigation procedures in assisting the elucidation of protein function at a molecular level.

## 1 Introduction

---

In this thesis, the allosteric regulation of the diguanylate cyclase PleD of *C. crescentus*, which synthesizes the bacterial second messenger c-di-GMP implicated in motility, biofilm formation and pathogenicity, is characterized by molecular dynamics simulations. In particular, PleD activity control via product inhibition is investigated by studying the liganded, c-di-GMP present in the inhibition site, and unliganded protein by means of structure and dynamics and results thereof are reported in the publication "Schmid and Meuwly 2007" (section 5.2). Biochemical studies discovered an additional level of PleD control, which is phosphorylation-mediated dimerization. Computational methods are applied to probe the sequence of events in the dimerization process, i.e. the local conformational loop rearrangement characteristic to CheY-like receiver domains, reorientation of the domains and the final tightening of the dimerization interface.

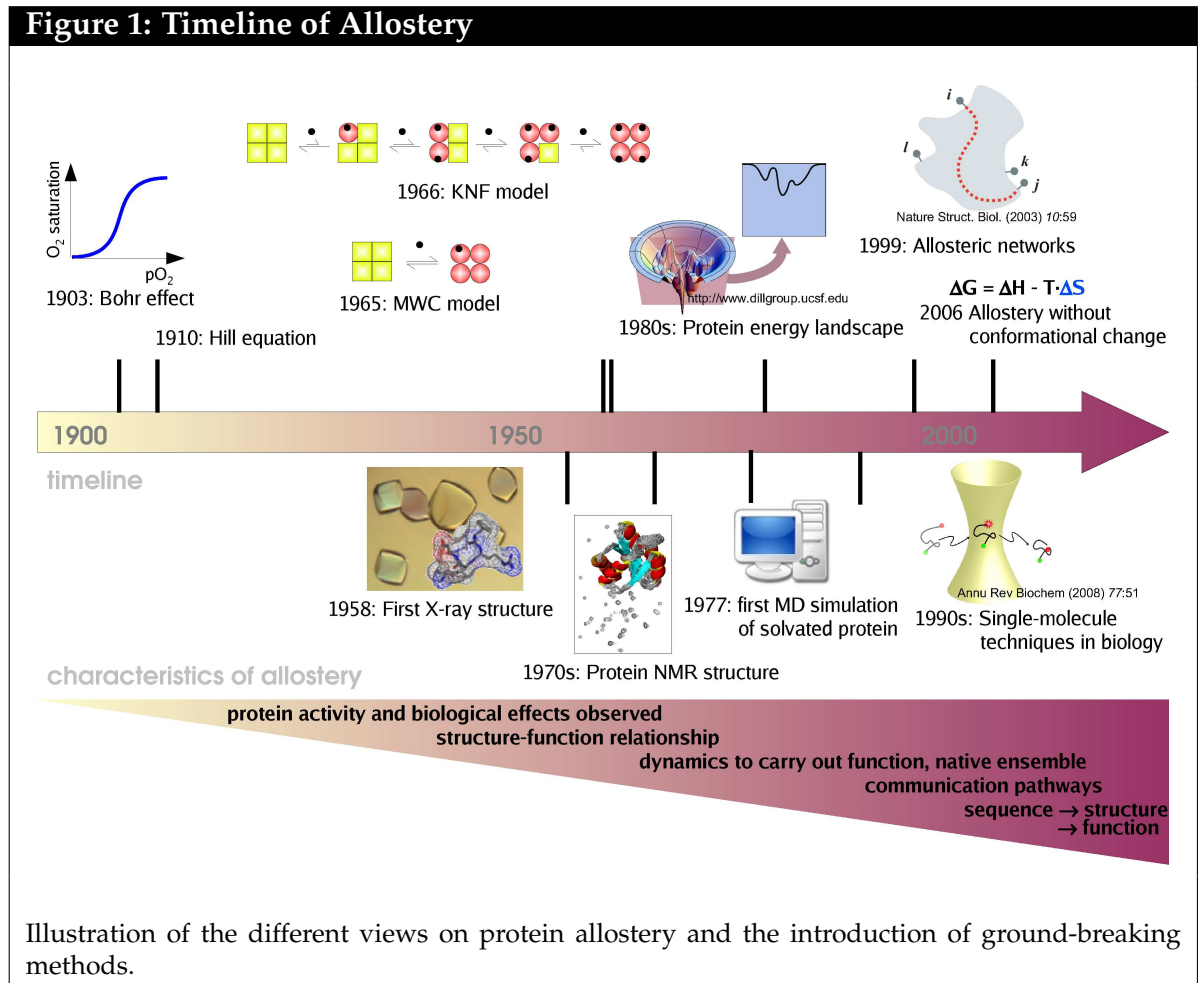
## 2 Protein Allostery

Allostery describes the phenomenon that a change at one site, the allosteric site, affects the activity at another, and in turn controls processes such as signal transmission, catalysis and turning genes on or off. Various processes such as binding of ligands, modifications via phosphorylation or glycosilation, change in pH, may trigger allostery. It is fascinating how nature propagates information from the allosteric site to functional parts of the molecule many tens of Ångströms away. Despite decades of research on allostery and the proposition of many mechanisms and models, presented in the following, the detailed molecular characteristic underlying the communication between distant sites is still not completely understood (see Figure 1).

### Structure-function paradigm

Before three-dimensional structures of biomolecular systems became available, allostery was investigated by measuring binding/reaction rates. Methods were developed to explain the measured enzymatic reaction rates, e.g. the Hill coefficient<sup>1</sup> to quantify cooperativity or Michaelis-Menten kinetics assuming the enzyme to be non-allosteric. Indeed, the sigmoidal oxygen binding curve of hemoglobin was measured in 1904 by C. Bohr, long before the first structure was solved. With the determination of the liganded and unliganded structure of hemoglobin the concentration dependence of the O<sub>2</sub> binding affinity was explained by the existence of two quaternary structures, the low-affinity deoxy T-state and the high-affinity oxy R-state.<sup>2,3</sup> Since then X-ray crystallography and NMR methods have steadily improved, and nowadays high-resolution three-dimensional structures are available for approximately 50,000 proteins and still growing. This increasing wealth of structural data, especially if a protein is solved in different states of activation or with/without bound ligand(s), allows the elucidation of structural features characteristic of a specific state. In most allostery-regulated systems a change of shape between the two distinct states was observed visually and hence allostery became tightly associated with a structural change. Even today, protein function is often explained via structural changes, e.g.

## 2 Protein Allostery



open and closed state of a binding site to accommodate and release substrate, surface remodelling to allow protein-protein association and movements of domains with respect to each other. Allosteric conformational changes are most prominent in biological motors, i.e. proteins that convert chemical energy to mechanical energy such as ATP-ase or myosin, but the scale of structural changes ranges from tens of Ångströms to sub-Ångströms (see Table 1).

### Energy-landscape concept & population-shift model

The "structure-function" relationship states distinct key structures for different functional states. The two classical models put forward shortly after the structure determination of tetrameric hemoglobin are the following: The Koshland-Nemethy-Filmer (KNF) model<sup>4</sup> describes the structural transition between the T- and R-state as an induced-fit mechanism, where binding of a ligand causes a subsequent struc-

**Table 1: Amplitudes and timescales of protein motions**

	Function	Amplitude [Å]	Timescale	Methods
<b>Local motion</b> atomic fluctuation sidechain flips break/formation of salt-bridges	<b>flexibility to allow reactions, and ligand docking</b>	< 1	fs	MD simulations X-ray diffraction NMR relaxation Fluorescence Spectroscopy (UV-VIS, IR, Raman) H/D exchange
			ps	
			ns	
<b>Medium-scale motion</b> loop motion terminal-arm motion subunit motion	<b>active site adaptation, binding specificity</b>	1 - 5	μs	
			ms	
			s	
<b>Large-scale motion</b> domain motion subunit motion	<b>hinge-bending motion, allosteric transitions</b>	5 - 10	hrs	
<b>Global motion</b> folding/unfolding subunit association	<b>ligand binding</b>	> 10		

The complex process of allostery involves subtle sub-Ångström structural changes to large-scale domain motions and protein-ligand associations occurring on a timescale of femtoseconds to hours which requires many different techniques for detection.

tural change. The subunits of hemoglobin switch independently but cooperatively facilitate further transitions to the high-affinity state. In contrast, the main idea of the Monod-Wyman-Changeux (MWC) model<sup>5</sup> is the existence of different interconvertible states, e.g. all subunits of hemoglobin must exist in either the T- or R-state. Binding of a ligand leads to a shift in the thermal equilibrium towards one state. Support for the latter model comes from nuclear magnetic resonance (NMR) studies that demonstrate the co-existence of different conformations.<sup>6,7</sup> Even in the absence of an allosteric trigger both functional states are observed, although differently populated. Hence, ligand binding favors a pre-existing structure and the allosteric process can be viewed as a shift in the thermal equilibrium. The different co-existing conformations can be attributed to local minima on an energy landscape<sup>8,9</sup>, a model originally developed for protein folding, that are divided by energy barriers. The energy landscape is defined as the potential energy of the system as a function of all coordinates. In this concept, the allosteric process changes the energy landscape by either lowering the end state minima or lowering the energy barriers, thereby allowing transitions to other conformations.

### Role of dynamics

A complete description of proteins includes not only an ensemble of conformational states, but also information about the dynamics, i.e. the interchange of the populations.<sup>10,11</sup> Static structures representing different functional states are so prevalent that it is easy to forget that proteins are dynamic molecular machines and rates of ligand binding or release may be governed by conformational gating owing to the intramolecular dynamics. Advances in nuclear magnetic resonance (NMR) methods provided data on exchange rates, populations and differences in chemical shift.<sup>10,12-14</sup> In addition, real-time dynamics can be observed by single-molecule fluorescence resonance energy transfer (FRET)<sup>15</sup>, a "spectroscopic ruler" to monitor distances, and more recently using time-resolved X-ray experiments<sup>16-18</sup> which allows the occupation of intermediate states and the kinetics of conformational changes to be followed at atomic level. Such studies reveal that proteins are highly flexible, not only through movements of surface loops and sidechains, but also by collective motion of the core structure.

Since the energy landscape is an intrinsic property of a protein and determines the dynamics of the system, efforts are made to map out this complex energy-landscape which contains the necessary information about the conformation ensembles, energy barrier heights and transition pathways. In particular, computational approaches to complement experimental data can prove very useful in providing links between structure, energy and dynamics. The phenomenal increase in computer power, particularly through the development of massively parallel machines, allows the study of biomolecular systems in explicit solvent approaching microsecond timescales. In addition, methods exist to sample the conformational space<sup>19</sup>, to probe the intrinsic flexibility inherent in the shape of the molecule<sup>20-24</sup> and combine experimental NMR parameters with simulations<sup>7,25</sup>.

### Enthalpy vs. entropy

An allosteric event such as ligand binding changes the energy landscape and the subsequent change in free energy governs the equilibrium shift. The free energy involves enthalpic and entropic contributions: structural changes and protein-ligand interactions contribute to the enthalpic term whereas changes in protein flexibility and dynamics alter the system's entropy. Taking this into consideration, Tsai *et al.*<sup>26</sup> proposed three types of allosteric proteins based on the protein's main driving force:



First, proteins showing large structural rearrangements and being dominantly governed by enthalpy. Second, entropy driven systems where protein motion rather than structure is affected, and third proteins where both enthalpy and entropy contribute. There exist numerous examples for structural changes, evident from the acceptance of the "structure-function" relationship. Yet, a first example of a biological system where allostery is mediated exclusively by the change in protein motions was discovered.<sup>27</sup> Thus, allostery surely is not restricted to a "change in shape"<sup>28</sup> and it was even proposed that every protein has allosteric potential<sup>29</sup>.

### **How to get a signal from one site to the other?**

Knowledge of functional key structures, their interconversion rates and protein flexibility are accumulating. The next major challenge is to elucidate low-energy pathways connecting the different known stable conformations of a system and to identify motions or structural features crucial for protein function. It is believed that local fluctuations (ps to ns timescale) modulate global structural transitions ( $\mu$ s to ms timescale) and indeed, normal mode analysis (NMA) indicates that motions driving the allosteric conformational change already exist in the (crystallographic) end states.<sup>23,30-32</sup> Hence, analysis of these usually low-frequency modes assists the elucidation of functional motion. Collective motion of distant parts potentially indicates dynamical coupling of the remote sites and domain reorientations can often be described by a hinge or shear mechanism.

It is believed that pathways connecting conformational substates are not random but predefined and inherent to the protein because biological function is the result of evolutionary selection. Analysis of co-evolved residues<sup>33,34</sup> and tracking of the energy flow through the protein<sup>35,36</sup> highlight a communication network and argue for the existence of specific signal transduction pathways, a concept further corroborated by mutagenesis experiments.

Understanding how a signal is communicated through a protein is a major challenge in structural biology. Albeit crystallographic structures are silent about the nature of the transitions from one state to the other and the forces involved, increasing time-resolved data start to shed light on the dynamic processes. A complete insight into allostery requires not only to highlight structural changes, but also the knowledge of when and where they occur and whether these changes are correlated or sequential. A useful tool to detect direct causality and coupling in allosteric processes is to drive conformational changes and watch the protein structure respond to the pertur-

## 2 Protein Allostery

---

bation.<sup>37-40</sup>

Clearly, protein allostery has many different aspects<sup>41</sup> that are all to some extent observed in biological systems, and which need to be reconciled in a comprehensive model.<sup>37,42</sup> Yet, to fully understand allostery, a continuous information exchange between experimentalists and theoreticians is necessary. Based on combined data from simulations and experiment, mechanistic models can be proposed which in turn need verification by further experiments. In addition, advances and innovation of methods in different scientific fields will have a considerable impact.

### Views on protein allostery:

- functional states have different conformations, e.g. opening/closing of active site
- the signal propagates along a predefined communication network which is consistent with the sequential KNF model
- co-existing conformations on an energy landscape with ligand binding inducing a population-shift and being related to the concerted MWC model
- enthalpy or entropy driven protein allostery

## 3 Methods Survey

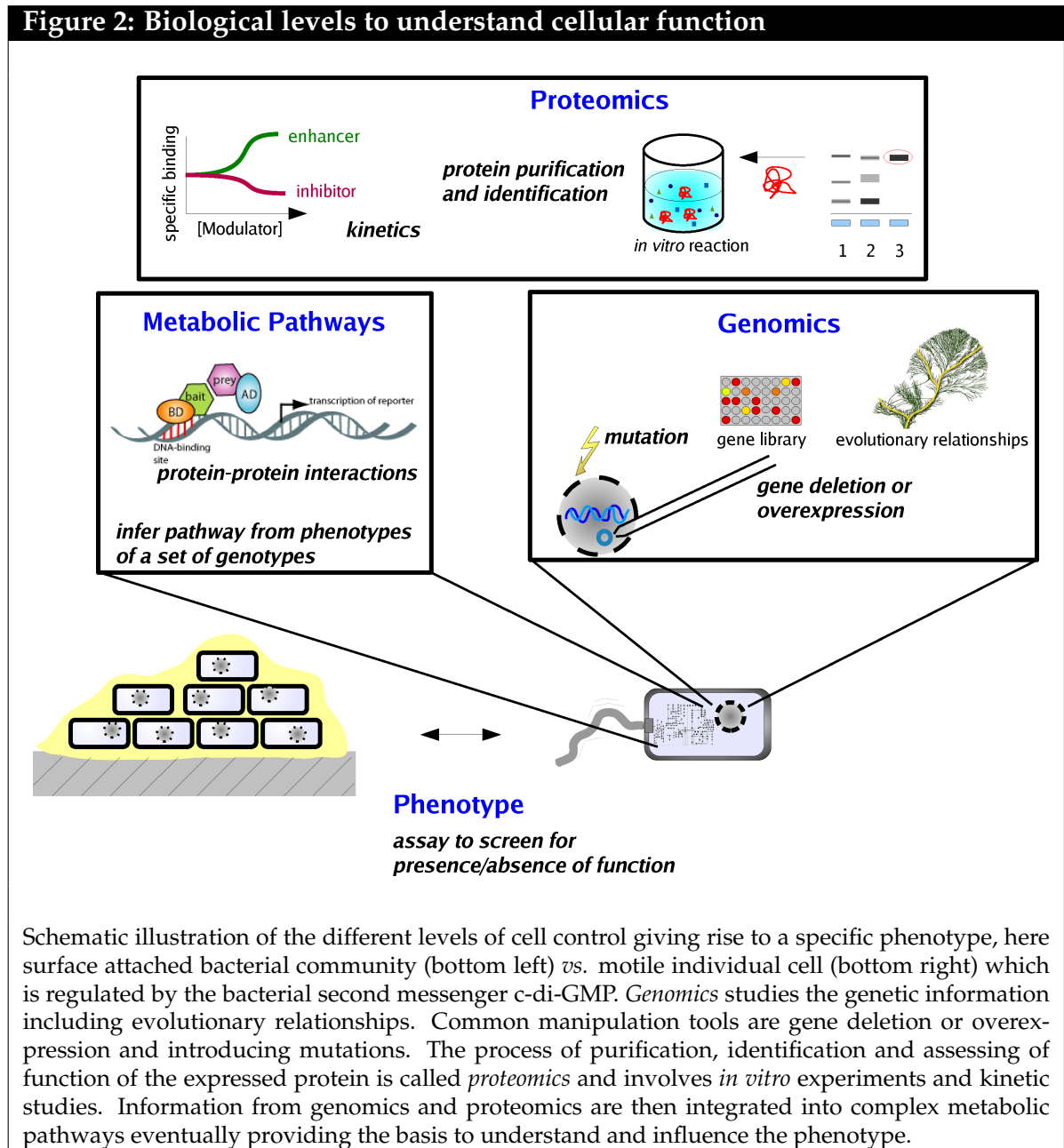
Methods were developed and/or adopted from the fields of biochemistry, physics and computer science to get insight into protein function, regulation and allosteric control. The many techniques collect complementary data, such as structure, kinetics, dynamics and pathways at different levels of resolution. Together they give a more complete picture of how biological systems work. The following survey presents common tools with emphasis on the contribution of computational methods. The list is far from being exhaustive, but should rather illustrate aspects of signal transfer from past successes.

### 3.1 Elucidating biological function and pathways

Allostery, a change in structure or dynamics induced by a modulator, is involved in the cellular response to signals and regulation of metabolic pathways. Biological function is the complex consequence of the action of a large number of molecules that interact in many different ways. Comprehensive characterization and identification of intracellular proteins, metabolites, and description of their interplay in biochemical pathways requires analysis at the transcriptome, proteome, metabolome, and phenotypic level (see Figure 2).

#### **Genomics**

Genomics includes the intensive efforts to determine the entire DNA sequence of organisms and the investigation of single genes, their functions and roles. Understanding of the latter is essential and involves genetic manipulation, e.g. gene deletion or gene overexpression, followed by the determination of changes in the phenotype and metabolism. Microarray technology is commonly used for genetic screens, a procedure to identify and select individuals who possess a phenotype of interest. Genome sequencing has showed that there exist conserved functional units and the study of



homologous proteins can help to find a common mechanism, a concept utilized in bioinformatics. Evolutionary relationships can also assist in finding interaction partners assuming two interacting proteins evolved by gene fusion and separation thereafter.<sup>43</sup> In the laboratory, genetic techniques are useful to produce a large amount of a specific protein and even allow to introduce point mutations in defined locations. Alanine scanning mutagenesis and site directed mutagenesis are key experimental tools to probe or confirm the importance of a specific residue involved in signal trans-

fer, binding or enzymatic activity. However, it is a very time-consuming process and suggestions coming from computational methods could narrow the search.

### Proteomics

Proteins are the main components of the physiological metabolic pathways of cells and proteomics is considered the next step in the study of biological systems, after genomics.

*Determine the existence of proteins in complex mixtures.* Cell lysis is followed by gel electrophoresis of the sample separating proteins according to their characteristics, such as charge or molecular weight. The distinct bands can be visualized with either un-specific dyes, e.g. Coomassie blue, or the protein of interest can be detected with antibodies specific to the target protein carrying a radioactive label or a reporter enzyme such as horseradish peroxidase. The latter approach, enzyme-linked immunosorbent assay (ELISA) is a common tool in biochemistry and cell biology studies and allows quantitative determination of protein amounts. In addition, chromatographic separation methods including gas chromatography (GC), high-performance liquid chromatography (HPLC) and capillary electrophoresis (CE) can be used. The presence of the target protein is then identified using mass spectrometry techniques.

*In vitro testing for protein functionality.* The study of enzyme kinetics can help us understand the function and regulation of enzymes. The most simple experiment is to measure the enzyme velocity at different substrate concentrations and fit the data to the Michaelis-Menten model assuming steady-state conditions.<sup>44</sup> The maximum enzyme velocity,  $v_{max}$ , and the concentration of substrate needed to get half-maximal velocity,  $K_m$ , are easily extracted from graphical representations such as the Lineweaver-Burk plot<sup>45</sup> or the Eadie-Hofstee diagram<sup>46</sup> or nowadays calculated more accurately by non-linear regression analysis. More importantly, the effect of small molecules and protein modifications such as phosphorylation or glycosylation on the enzymatic activity of a protein is readily detected by kinetics measurements. A modulator can act as activator or inhibitor and can bind to the same site as the substrate (competitive) or to another allosteric site (non-competitive, uncompetitive) and measurements at several concentrations of inhibitor help to distinguish between the models of enzyme inhibition. If the modulator is identical to the substrate the term feedback control is used. If the enzyme has cooperative subunits a sigmoidal curve is observed and the additional parameter  $h$ , the Hill coefficient, is introduced.<sup>1</sup> A value of  $h = 1$  corresponds to no cooperativity and is identical to the standard Michaelis-Menten equa-

### 3 Methods Survey

---

tion, while  $h > 1.0$  and  $h < 1.0$  correspond to positive and negative cooperativity where the binding of a molecule is facilitated or hindered.

Protein selection, identification and characterization of enzymatic properties enables the assignment of protein function. With the knowledge of substrate and product compounds as well as allosteric mediator molecules the protein can be integrated into a metabolic network.

#### **Protein-protein interactions**

Most proteins only function in collaboration with other proteins, RNA molecules or other small ligands. They may even require protein modifications to become a target for binding or interacting with a distinct set of other proteins. It is especially useful to determine potential partners in cell signaling cascades, and methods to detect and analyze protein-protein interactions have been reviewed.<sup>47</sup>

The *yeast two-hybrid screen* is a traditional method to detect protein-protein interactions and was originally developed by Fields and Song in 1989.<sup>48</sup> In brief, the transcription factor is split into two separate fragments, the binding domain and the activating domain. A "bait" protein to be tested for interaction is fused to a DNA-binding domain, while a library of proteins (the "prey") are fused to the activation domain and transformed into yeast. An interaction between "bait" and "prey" brings the two fragments in close proximity and initiates transcription of a reporter gene. Utilizing genetically engineered strands that allow positive or negative selection, i.e. cells with successful protein interactions live or die on the provided media, allows easy read-out. After this selection the proteins displaying the appropriate characteristics need to be determined. This method has since been extended to detect DNA/RNA-binding proteins and molecules that disrupt protein interactions.

In *crosslinking experiments*<sup>49</sup> a covalent bond is introduced and transient interacting partners can be identified. Exposure to UV light causes an incorporated photoreactive group to covalently link to nearby molecules, thus "freezing" in place any interacting protein as a complex. Subsequent digestion and mass spectrometric detection of linked fragments identifies the position of binding and assists in the determination of multimeric states.

Protein-protein interactions are of central importance for virtually every process in a living cell. To study associations and elucidate signal transduction networks, high-throughput screening techniques were developed. Recent advances in microscopy have begun to visualize their dynamic behavior and the localization of specific pro-

teins within the cell.

### Phenotype determination and establishing biochemical pathways

The phenotype is any detectable characteristic of an organism which is determined by its genotype and environment e.g. morphology, development or physiological properties. To detect changes in the cellular response coming from genetic manipulation, assays have to be designed to screen for the presence or absence of the function investigated. Finally, combined gene and protein information must be correlated with particular phenotypic settings and a network of metabolic reactions can be devised, though it is a difficult task because the outputs of one enzymatic chemical reaction are inputs to other chemical reactions, creating very complex cellular signaling pathways.

All three levels - genomics, proteomics and phenotype - are interweaved and required to draw conclusions. In the case of the bacterial signaling molecule c-di-GMP which controls motility and biofilm formation, its presence or absence is easily tested by the determination of the morphology of the colony, the attachment to surfaces and cellulose production. The protein domains synthesizing and degrading the compound were confirmed by *in vitro* studies, and genome screens for these domains discovered the ubiquity in bacteria. Detailed information on the function and regulation of the second messenger c-di-GMP is given in section 4.

#### Biochemical and genetics techniques

- identify the molecule of interest, e.g. protein or signaling compound, determining a specific phenotype
- detect interacting partners and effects on functionality when bound
- in particular evolutionary relationships assist in elucidating common regulation mechanisms

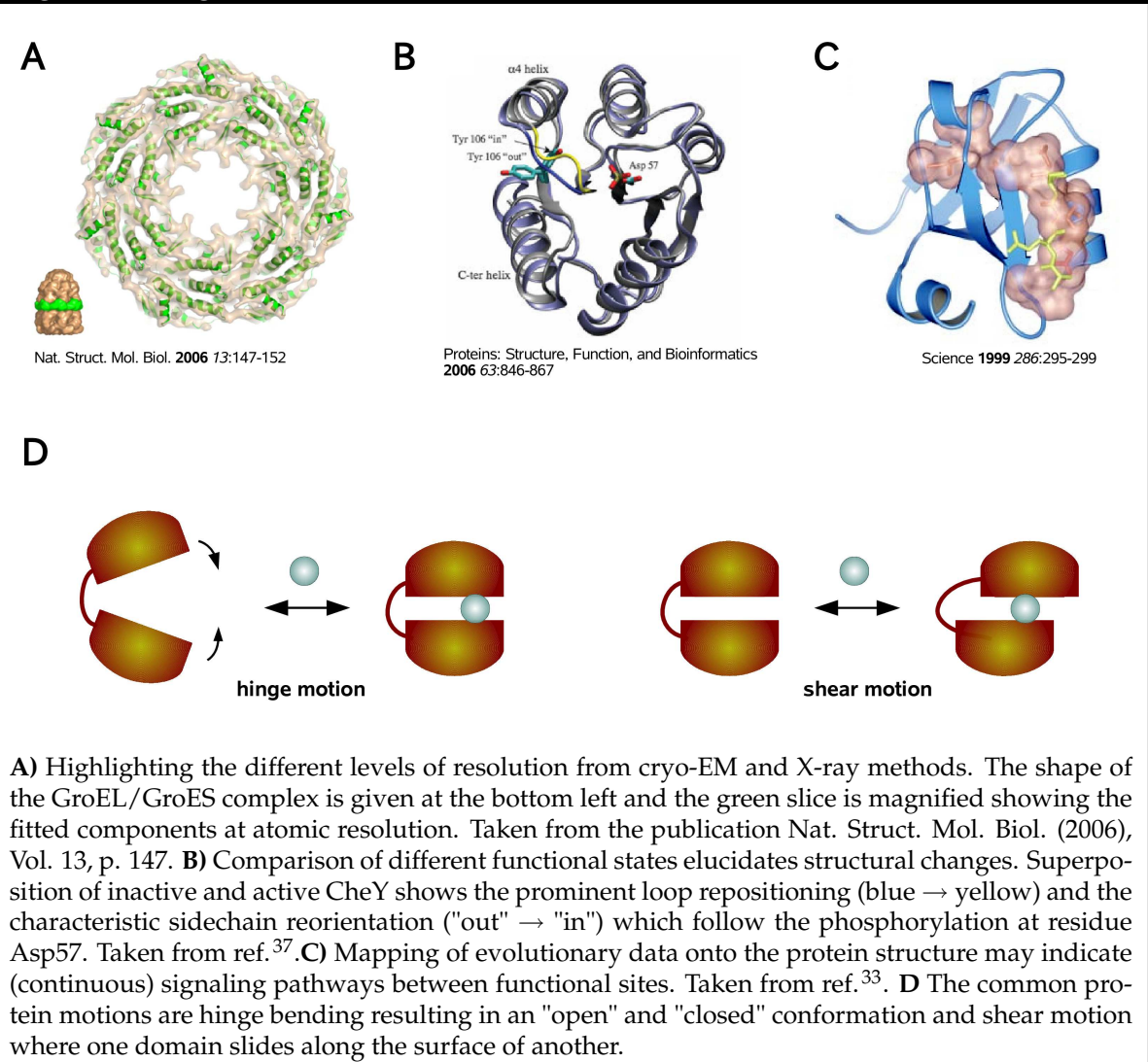
## 3.2 Structure determination

Biochemical and genetics techniques determine the interaction partner(s) of a given protein, highlight the ultimate cellular characteristics known as phenotype, and attempt to describe the inter-relations of functional modules in biochemical pathways.

### 3 Methods Survey

However, the knowledge that a molecule binds to a protein and thereby modulates its function gives no information on how the protein exerts function or what mechanisms are involved in allosteric control. Thus, three-dimensional (3D) structural data is invaluable and has contributed tremendously to our understanding of protein function at an atomic level.

**Figure 3: Insight from Structure Determination**





### 3.2.1 X-ray crystallography

#### Obtaining high-resolution structures

X-ray crystallography is the traditional method to elucidate structure at high resolution ( $\leq 3$  Å). A beam of X-rays is directed at a crystalline sample and the diffraction pattern is then interpreted to produce a 3D model of the protein. Using this method, the structures of numerous proteins and complexes, e.g. protein-protein, protein-small molecules, protein-DNA/RNA were solved, and it has become the major source of structural information (85% of all PDB entries). However, it is limited to molecules that form regular crystals, hence structural information on membrane proteins and flexible fibril-like proteins is scarce. Although structural features may be distorted by unphysiological crystallization conditions or lattice contacts, X-ray crystallography proved useful in identifying structural changes such as large-scale domain motions or more subtle loop and sidechain reorientations that are related to protein function. Web-based programs offer analysis and visualization of potentially functional motions for two protein conformations provided (see section 3.5.6).

CheY is the response regulator of bacterial chemotaxis and serves as a model system for receiver domains. Comparison of two functional states of CheY – an inactive and  $\text{BeF}_3^-$ -activated state – first revealed repositioning of the  $\beta 4$ - $\alpha 4$  loop together with the coupled sidechain movement of residues T87 and Y106 which is now a commonly accepted characteristic of the activation process.<sup>50</sup> These changes are subtle and the RMSD is of a few Ångströms only. Although key structural changes were identified their exact role in the allosteric process is still under investigation and requires input from further methods. Another example where structural knowledge assisted in establishing a mechanistic model of function is the diguanylate cyclase PleD.<sup>51,52</sup> Allosteric product inhibition of PleD involves cross-linking of two domains by the ligand thereby frustrating domain flexibility and blocking productive encounter of two active sites (the model is depicted in Figure 6 B). Moreover, the structure allowed localization and characterization of ligand binding.

#### Watching a protein function

In general, X-ray crystallography provides static snapshots of the system. Time-independent acquisition of diffraction data of short-lived intermediates can be obtained by chemical trapping or freeze-trapping, where molecular motion is frozen by substantially lowering the temperature. In this case the time resolution is defined by the offset between the reaction initiation and the flash-cooling conserving the current

conformation. Advancement to true time-dependent crystallography, where pump-probe methods are used and the time progress is studied, leaving very short time for data collection, required various technical efforts.<sup>53,54</sup> First, triggering of the studied reaction must be very fast (fs to ns pulse) and uniform throughout the entire crystal, e.g. adding substrate or photoactivation. Second, data acquisition time must be reduced whilst maintaining high completeness in covering the reciprocal space. And third, sampled intermediate structures must be extracted from the time-dependent data.

Although still in its early days, the technique of time-resolved protein crystallography, i.e. the capturing of a sequence of pictures allowing to "watch" the protein as it executes its function, has demonstrated its capability to detect quite subtle but essential structural changes in myoglobin following the ligand photodissociation.<sup>16,17</sup> Recently, the allosteric R-to-T transition upon ligand dissociation in dimeric hemoglobin of the mollusk *Scapharca inaequivalvis* was followed with time-resolved X-ray.<sup>18</sup> On the timescale of microseconds, a two-step mechanism was observed: Structural changes in the heme region, rearrangements in the neighboring F-helix and CD-loop region as well as disruption of the water molecule network facilitate the sidechain reorientation of residue F97 and the subsequent structural transition. Hence, time-resolved crystallography clearly provides the unique opportunity to obtain high-resolution structure information as the protein undergoes structural changes. The conformational changes must, however, be accommodated within the crystal.

#### 3.2.2 Cryo-electron microscopy

Cryo-electron microscopy (EM)<sup>55,56</sup> uses a transmission electron microscope to examine frozen hydrated samples providing low-resolution structures ( $\geq 7 \text{ \AA}$ ) of large complexes and agglomerates. The 3D pictures indicate the overall shape and are best suited for analysis of secondary, tertiary and quaternary structure, the levels at which structure-activity relationships operate. The possibility to image the shape of large assemblies renders cryo-EM a complementary tool to X-ray and NMR. A most promising approach to structural analysis is then to combine the high-resolution with the low-resolution data. Atomic resolution structures of individual components are fitted into the EM density maps offering a more complete picture at various levels of detail.<sup>57</sup>

The now routine collection of 3D structural information mainly using X-ray, NMR and more recently cryo-EM remains essential to characterize potentially important structural features and binding modes leading to mechanistic proposals for protein function. Furthermore, biochemical and genetics information such as mutations or residue conservation can be mapped onto the protein structure, establishing the structure-function relationship. Atomic structures then provide a basis for computational simulations. Additional information on time-dependent changes in structure and the observation of transient intermediates will help to clarify protein mechanisms, especially pathways between resolved structures.

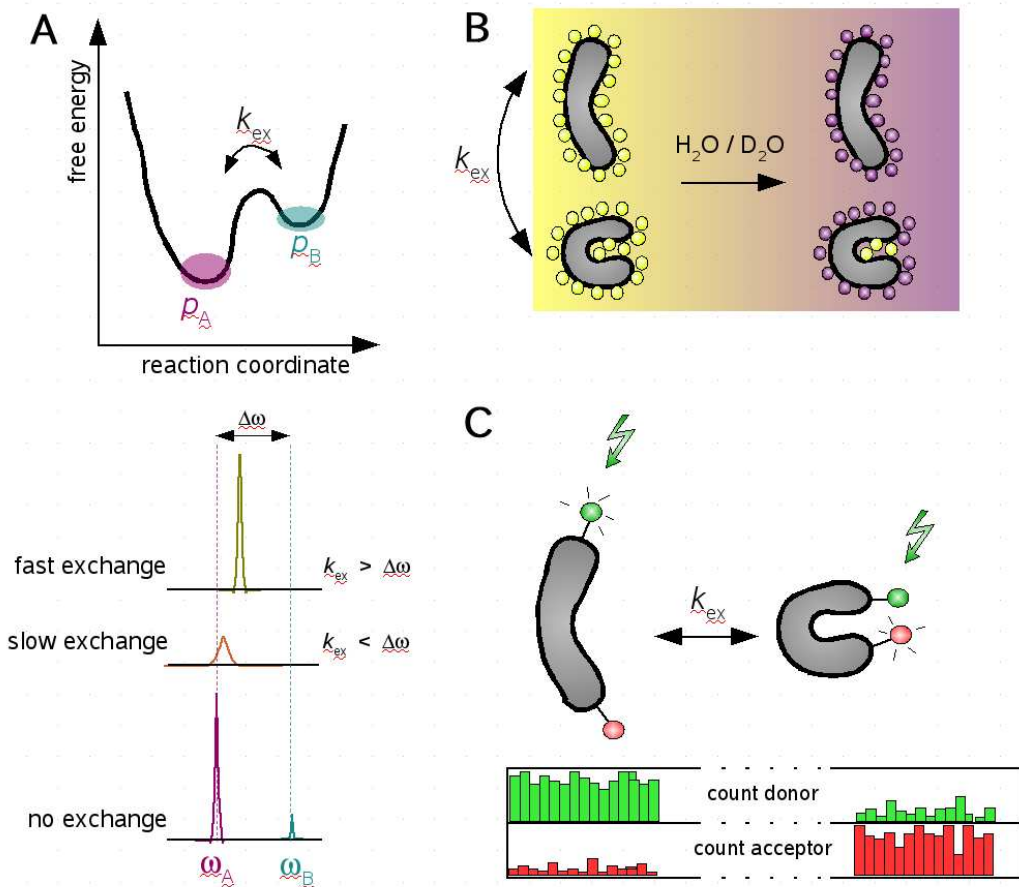
### Based on the three-dimensional structure

- mechanistic models are proposed from comparison of different functional states,
- interaction sites and binding modes are identified and characterized,
- biochemical and genetics data can be integrated to study structure-function relationships.
- very recently, time-resolved X-ray provides motion picture of processes.

## 3.3 Protein dynamics

With more and more structures of proteins known with high accuracy, one can easily forget that proteins are dynamic machines and flex to function. Many proteins undergo thermally driven transitions, so-called conformational changes, between two or more equilibrium structures: The transitions have important functional consequences for binding affinities and switching of regulatory proteins. Yet, little is known about the dynamics of these macromolecules, although growing evidence for the population-shift model stems from nuclear magnetic resonance (NMR) spectroscopy studies. In the following, advances in NMR to characterize dynamic processes, and fluorescence-resonance energy transfer (FRET) visualizing motion in real-time are reported.

Figure 4: "Watching" proteins move



**A)** The exchange between two conformational states A and B can be characterized with NMR, yielding thermodynamics, kinetics and structural information simultaneously. If they do not exchange, their relative population is given directly by the peak integral, and the chemical shifts provide structural data. The rate of interconversion,  $k_{ex}$ , can be deduced from the lineshape distinguishing fast and slow exchange. **B)** Hydrogen/deuterium exchange is rapid for solvent exposed sites and detects changes in the solvent exposed surface. The time scale of the protein motion observed is given by the time between quenching and analysis, often ranging from subseconds to hours. **C)** Fluorescence-resonance energy transfer (FRET) spectroscopy provides real-time data of distances between two labeled sites. However, single-molecule resolution is required and limits the observation time period.

### 3.3.1 Nuclear magnetic resonance (NMR) spectroscopy

Initially, nuclear magnetic resonance (NMR) experiments, a spectroscopic method that takes advantage of the nucleus magnetic moment by alignment with the electromagnetic field, were applied to elucidate the solution structure of small proteins up to

100 kDa. The time-consuming crystallization process required for X-ray crystallography is avoided, but the structure needs to be deduced from chemical shifts describing the local structure and nuclear Overhauser effect (NOE) data measuring short interatomic distances. However, NMR can be used beyond pure structural description offering site-specific information about protein motions over a very wide range of timescales, from picoseconds to hours. Recent advances in NMR relaxation methods and chemical exchange measurements allow characterization of both fast, picosecond to nanosecond, and slow, millisecond to microsecond, protein motion.<sup>12-14,58</sup> Slow motions are of considerable interest to allosteric transitions and ligand binding, whereas fast motions have an effect on the entropy of the system.

### **NMR spin relaxation is an indicator of internal motion**

Relaxation rates, i.e. the time needed for excited nuclear spins to return to their original equilibrium state, contain information about intramolecular dynamics at specific atomic sites. Specialized relaxation dispersion experiments have been developed, allowing determination of exchange rates between conformational states ( $k_{ex}$ ), their population ( $p_A$  and  $p_B$ ) and backbone dynamics.<sup>14,58</sup> The dynamics information is encoded in the intensities or lineshapes of the resonance signals, shown in Figure 4 A, but unfortunately does not provide direct structural information about the different conformational states. To simultaneously obtain structure and dynamics, experimental structural information from NOEs together with NMR relaxation data are used as restraints in ensemble-averaged simulations.<sup>7,25</sup> Thus, computational methods provide a means to link these spatial and dynamic characteristics. This approach successfully demonstrated that the opening movement of the two EF hands in each domain is coupled *via* correlated motion of the helices. Structural fluctuations in free calmodulin overlap with complex-like substates.<sup>7</sup>

*Two distinct populations that interchange.* NMR results strongly support the population-shift model, the co-existence of two different states with redistribution of their population during the allosteric event.<sup>6,7</sup> If the exchange rate,  $k_{ex}$ , is sufficiently slow then resonances from both states are observed and can be analyzed in terms of structure. The relative population,  $p_A$  to  $p_B$ , of exchanging species is given by the ratio of the peak integrals and provides thermodynamics information. As the exchange becomes faster, the two individual peaks disappear and a single averaged resonance line is observed at the population-weighted average shift.

*Identification of binding sites.* Differences in <sup>15</sup>N relaxation dispersion values reflect

variations in the chemical shift between the two protein states, and can identify residues which are affected directly by ligand binding. This equilibrium perturbation NMR method proved effective in the regulatory unit of protein kinase A to identify the known phosphate binding cassette region and hot-spot sites found previously.<sup>59</sup> In addition, in paramagnetic NMR relaxation enhancement effects (PRE), an unpaired electron increases the relaxation rate of other nuclei in an  $r^{-6}$  distance-dependent manner, and will detect all residues within a 15 – 35 Å radius of the spin label.<sup>60</sup> Recent work involves site-directed spin labels in which the label has been inserted into proteins at a specific position to confirm, for example, conformational changes<sup>30</sup>.

#### **Long-range information from residual dipolar couplings (RDC)**

A traditional weakness of NMR is that all the structural restraints are short-range in terms of distance, e.g. NOE restraints are only between atoms  $< 5$  Å apart. For large macromolecules the uncertainties will add up and the overall structure will be poor even though individual regions of the structure are well defined. Residual dipolar coupling (RDC) measurements of macromolecules that are partially aligned in dilute liquid crystal solution, preventing complete averaging of the dipolar interaction, yield long-range angular information. This method proved useful to accurately define the relative orientation of individual parts with known structure in the overall system, e.g. domains, secondary structure or subunits of complexes, and the graphical rigid-body modelling program "Module" facilitates the reorientation of user-defined protein parts into a common display frame.<sup>61</sup>

*Slow correlated motion.* In addition to providing global structural information, RDCs also comprise dynamics data, specifically at slow timescales in the submicro to millisecond range at which allosteric processes occur. The 3D-Gaussian axial fluctuation (GAF) model was used to interpret the dynamics of the peptide plane in the Ig-binding domain of streptococcal protein G.<sup>62</sup> A long-range network of slow correlated motion transmitted across interstrand hydrogen-bonds was found. This propagation of slow motion across the entire  $\beta$ -sheet clearly has implications for understanding collective motions and long-range information transfer in proteins in general. Furthermore, the ability to measure scalar couplings across hydrogen bonds<sup>63</sup> allows unambiguous identification of the donor and acceptor atom, measures the strength of interaction and can even be related to geometry<sup>64</sup>.

**H/D exchange studies to observe motions on the subsecond to hours timescale**

Hydrogen-deuterium (H/D) exchange is a chemical reaction in which a covalently bonded hydrogen atom is replaced by a deuterium atom when rapidly changing the solvent from  $\text{H}_2\text{O}$  to  $\text{D}_2\text{O}$ , or vice versa (illustrated in Figure 4). Usually the examined protons belong to the amides in the protein backbone. Based on the assumption that the exchange rate reflects the exposure of a particular amide to the solvent, changes in local dynamics are observed from changes in the exchange rate, e.g. upon partial unfolding or conformational changes altering the solvent accessible surface. H/D exchange is followed using NMR or mass spectrometry and the time scale of milliseconds to hours, inaccessible using relaxation methods, is given by the elapsed time between initiation and analysis. In NMR the disappearance or appearance of signals is observed by measuring multiple spectra. It is a powerful method not only to track slow conformational changes, but also protein-protein interactions and protein folding.

NMR spectroscopy is an exceptionally versatile technique having exquisite time and spatial resolution, and providing the time scale of transitions, the population of the different states and the atomic resolution structure. NMR confirmed the coexistence of different functional states underpinning the protein energy landscape and population-shift model to explain allostery, and is also valuable to map binding sites. However, raw experimental data needs to be combined with additional information for interpretation. Often structure ensembles that satisfy the experimental NMR restraints are generated with molecular dynamics simulations<sup>7,25</sup>, and NMR data can serve to improve force fields<sup>65,66</sup>.

**3.3.2 Fluorescence-resonance energy transfer (FRET)**

Spectroscopy measures the interaction between light and matter, i.e. absorbance and emission. Besides time-resolved infrared (IR) and Raman spectroscopy, fluorescence-resonance energy transfer (FRET) experiments prove very promising to discover protein motions, particularly in view of single-molecule resolution.

**Single-molecule techniques show details of individual molecules**

Recently, new techniques have been developed to measure and manipulate single biomolecules.<sup>67</sup> Specific particles can be identified and tracked, and protein interac-

tions or conformational changes can be observed which lead to new insights. Studying a single molecule, information can be obtained that otherwise would disappear in the ensemble-average, e.g. two conformations such as open and closed show different distances but average out to a medium length in the ensemble. Moreover, it may help validate models from molecular dynamics simulations, and even ensemble properties can be reconstructed from many single-molecule observations using occurrence-histograms.

#### A "spectroscopic ruler"

Fluorescence-resonance energy transfer (FRET) is based on the observation by T. Förster in 1948 that energy is transferred non-radiatively from an excited chromophore (the donor,  $D$ ) to another chromophore (the acceptor,  $A$ ) by means of intermolecular long-range dipole-dipole coupling.<sup>68</sup> The efficiency of energy transfer ( $E$ ) strongly depends on the sixth power of the distance  $r$  between  $D$  and  $A$ ,

$$E = \frac{1}{1 + \left(\frac{r}{R_0}\right)^6} = \frac{\text{photons transferred from } D \text{ to } A}{\text{photons absorbed by } D} \quad (3.1)$$

where  $R_0$  is the Förster radius an intrinsic property of the fluorophore. Hence, FRET is a sensitive reporter of distances in the range of 10-100 Å and often quoted as a "spectroscopic ruler" (see Figure 4 C). Improvements in site-specific attachment of fluorescent molecules, in addition to the use of the intrinsic fluorophores Tyrosine and Tryptophane, allow the measurement of specific spatial distances that are suggested to change during protein function or allosteric processes. Suitable FRET pairs are chosen based on their overlap integral, excitation/fluorescence wavelength and the distance of interest, being around the Förster distance  $R_0$ , to have maximum sensitivity. After labeling, the functionality and proper structure of biomolecules must be confirmed because bulky fluorophores might disturb the system. The observation time of single-molecule fluorescence is in the range of milliseconds, and is limited by photobleaching processes and the diffusion time through the observation volume if no immobilization is used.  $E$  is generally measured using ratiometric intensity measurements characterizing pair distances over time (see Equation 3.1) or fluorescence lifetime measurements. FRET is a popular tool for accurate measuring of single-molecule distances applied to monitor protein movement, complex formation and to detect enzymatic turnover.<sup>69,70</sup> If the timescale of interconversion between an open and closed conformation is of the order of the residence time of the molecule in focus or slower,



two separate  $E$  distributions are observed. One example for real-time observation of conformational motion is the demonstration of the three-step rotary movement of the  $\gamma$  subunit in  $F_0F_1$ -ATP synthase during catalysis.<sup>15</sup> The detailed topological information provided from FRET confirmed the opposite direction of  $\gamma$  rotation during ATP synthesis and hydrolysis, respectively. In addition, the duration of the FRET states were shown to correspond to the timescale of the catalytic event. Furthermore, labeling of each interaction partner will give information on complex formation and their respective orientation. In FRET-based bioassays enzymatic cleavage releases the donor and acceptor moiety separately into the solution, and the donor fluorescence is no longer quenched. Fluorescence lifetime imaging microscopy (FLIM) provides the localization of labeled particles within a cell, e.g. the membrane or the nucleus, which is of particular importance in *in vitro* studies.

FRET provides accurate spatial information over the course of time and has become a very useful tool of structural biology, indeed. Due to the fact that distances are measured, structure is needed to relate them to corresponding conformational changes. The remaining challenge is to find appropriate labeling sites that characterize functional motion in proteins.

#### Protein dynamics

- NMR yields
  1. populations together with their structure
  2. timescale of exchanging states
  3. identifies residues involved in binding molecules that show structural or dynamics changes
- single-molecule techniques detect characteristics of individual molecules otherwise hidden in the population average
- FRET detects distance changes in real-time

### 3.4 Thermodynamics

#### 3.4.1 Isothermal titration calorimetry (ITC)

Calorimetry measures the heat of chemical reactions or physical changes and is often used to study the thermodynamics of ligand binding. Isothermal titration calorimetry (ITC) directly determines binding affinity ( $K_d$ ), enthalpy changes ( $\Delta H$ ) and binding stoichiometry ( $n$ ). The association free energy  $\Delta G$  and the entropic component can thereby be calculated using the formula

$$\Delta G = -RT\ln K = \Delta H - T\Delta S. \quad (3.2)$$

Known amounts of ligand are titrated into the sample cell causing a temperature in-/decrease in the case of exothermic or endothermic reactions, respectively. The heat needed to maintain constant temperature is usually plotted as a function of time showing spikes at every ligand injection. The calculated total heat as a function of the molar ligand-protein ratio eventually yields thermodynamic information about the binding process. Knowledge about the enthalpic and entropic components to binding affinity became particularly relevant in view of protein dynamics governing allostery.<sup>27</sup>

#### Thermodynamics

- yields information about the total free energy change ( $\Delta G$ ) which is related to the binding affinity
- extracts separate contribution from enthalpy ( $\Delta H$ ) and entropy ( $\Delta S$ ), i.e. structure and dynamics of the biomolecular system

### 3.5 Computational contributions to mechanistic understanding

Computational approaches serve as a link between structure and dynamics.<sup>71-73</sup> The widely used classical molecular dynamics (MD) simulations start from a given atomistic structure, provided by X-ray crystallography, NMR spectroscopy or homology

modeling, and numerically calculate the motion of each atom according to Newton's equations of motion. Atoms are assigned van der Waals radii and atomic partial charges. The interaction potential between the atoms is given by a so-called "force field", which contains bonded and non-bonded terms (electrostatic and van der Waals forces). The level of accuracy of the force field depends on the question being addressed and the CPU resources available. Besides classical empirical models<sup>74</sup>, force fields containing quantum mechanics suitable to simulate chemical reactions<sup>75,76</sup>, polarizability<sup>77</sup> and more recently reactive force fields<sup>78,79</sup> were developed.

The phenomenal advances in computer technology in the last decades, especially high-performance parallel computing, has opened the possibility to investigate large protein systems in explicit solvent for timescales reaching microseconds, and culminated in the simulation of an entire virus.<sup>80</sup> However, following full atomistic motions in biomolecules in the time range of microsecond to seconds at which interconversion of allosteric subconformations take place is still not feasible with conventional simulations. Methods have been devised to tackle the time limitations, for example coarse-graining, the use of multiple-timescales, normal mode analysis, steered/targeted MD or Monte-Carlo sampling.

Computational approaches not only provide direct insight into protein dynamics at an atomic level, but also give energetic information. Atomic motions are easily visualized, and following the propagation of structural changes often results in the proposal of mechanistic models for protein function, highlighting the role of various structural motifs. Lately, the transition between known structural states is increasingly addressed in terms of causality, i.e. a sequence of events, and changes in protein flexibility. In addition, it is relatively straightforward to introduce mutations, to generate and characterize short-lived intermediates or to perturb the system and monitor subsequent changes, the latter being important to study signal transfer as in allostery. Computational methods are complementary to experiments insofar as potential mechanistic models based on simulations need verification with experimental data, while computations can assist in interpreting these data.

### 3.5.1 Root-mean square deviation (RMSD)

The root-mean square deviation (RMSD) is a quantitative measure of the distance between two superimposed structures, for example X-ray structures of different conformations.

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N (r_i - r_i^0)^2} \quad (3.3)$$

$N$  is the total number of atoms and  $r_i$  is the current position of atom  $i$  while  $r_i^0$  denotes the reference position. In MD simulations, the RMSD with respect to the initial starting structure is often plotted as a function of time, and is used to assess the extent of structural relaxation in the equilibration phase and to ensure the stability of the system throughout the entire simulation. Generally, the RMSD for backbone atoms is of the order of a few Ångströms. Large distortions from the experimental structure may be prevented by energy minimization using position constraints prior to molecular dynamics simulations. Prominent structural changes during the trajectory are readily observed as a sudden in- or decrease in RMSD simultaneously indicating the timescale.

Before calculating the RMSD the two structures need to be aligned with care to remove contributions from translation and rotation of the entire molecule. In multidomain proteins this process is especially delicate. Reorientation of domains with respect to each other can be detected by fitting on one domain but calculating the RMSD for the entire protein. Corresponding hinge axes can be identified using additional methods (details in section 3.5.6). One possible drawback of calculating RMSDs is the fact that it is an overall property, i.e. the same overall RMSD is obtained for rearrangement of "rigid-like" domains and structural changes within the domains without affecting their global orientation.

#### 3.5.2 Root-mean square fluctuation (RMSF)

Root mean-square fluctuations (RMSF) correspond to the time-averaged fluctuations of the atom position around a reference position, e.g. average structure or initial structure, and characterize the atom's mobility. Superposition of all snapshots is necessary to exclude rotational and translational motion of the molecule occurring during the simulation, thereby restricting analysis to internal motions of the molecule only. Calculated RMSF values can be related to  $B$ -factors from X-ray experiments:

$$B\text{-factor} = \frac{8\pi^2}{3} \cdot RMSF^2 \quad (3.4)$$

RMSF and  $B$ -factors are a measure of flexibility, e.g. regions of high RMSF values

correspond to very flexible regions, and agreement between the pattern of flexible and rigid regions obtained from theory and experiment corroborates the usefulness of MD simulations. However, it should be noted that crystal contacts have a potential influence. Comparing the RMSF profile for free and ligand-bound proteins can indicate changes in protein dynamics that may be related to protein function such as providing the necessary flexibility to orient residues allowing enzyme reactions or ligand accommodation, and having entropic consequences.

### 3.5.3 Dynamic cross-correlated motion

Analysis of correlated motion is important to highlight regions of the protein that move in a concerted manner, that is sites which are motionally coupled and possibly could communicate signals.<sup>81,82</sup> Correlation maintained during simulations denotes contact between structural parts and that a change in one site is reflected by residues at another. After removing rotational and translational motion by superpositioning each snapshot, the cross-correlation coefficient  $C_{ij}$  between two atoms  $i$  and  $j$  is defined as

$$C_{ij} = \frac{\langle \Delta r_i \Delta r_j \rangle}{\sqrt{\langle \Delta r_i \rangle^2 \cdot \langle \Delta r_j \rangle^2}} \quad (3.5)$$

where  $\Delta r_i$  is the displacement from the mean position of atom  $i$ , for a specific period of time. The trajectory is usually divided into successive blocks to examine the time-dependence of the correlation,  $C_{ij}$  is determined for each block and then averaged. The value of  $C_{ij}$  ranges from +1 to -1, for fully correlated and anti-correlated motions, respectively. If the motions of atoms  $i$  and  $j$  are random with respect to each other,  $C_{ij}$  is equal to 0. Commonly, cross-correlation matrices, also called dynamic cross-correlation maps (DCCM), are used to visualize individual residues or secondary structure elements that move in a concerted manner. In contrast to the detection of long-range couplings implying distant communication, coupled motion due to secondary structure, i.e. spatial or sequence proximity, present as broadening of the diagonal for  $\alpha$ -helices and off-diagonal for  $\beta$ -sheets is less interesting. Unfortunately, correlation coefficients do not show any information about the magnitude or the direction of the motion.

#### Elucidating signaling pathways based on correlated motion

Cross-correlated motion analysis was applied successfully to the elucidation of signal transduction in immunoglobulin G.<sup>82</sup> A small  $\beta$ -sheet (NC sheet) was identified as the key structure to transmit conformational changes between domain V and C. A further example is the observation of ligand induced changes of correlated motion in PleD detected even at long distances from the binding site.<sup>81</sup> Enhancement of the coupling between domains D2 and DGC is easily explained by the presence of c-di-GMP linking the two domains. In contrast, stronger correlation between the remote domains D1 and D2 in free PleD seems to be allostery based and potentially influences the dimerization behavior. From a more detailed analysis, two communication pathways connecting correlated motion of the inhibition and active site can be envisaged.

#### Identification of compact entities

Correlation maps can outline movements of protein parts that are strongly correlated, forming a "rigid" entity, and are often used in conjunction with normal mode analysis. Low-frequency normal modes describe large-scale motions which are likely functionally important motions, and it is desirable to dissect these into structural motifs moving concomitantly in the same or opposite direction. Even so, careful analysis is vital and one should note that coupling of structural motifs in a small number of low-frequency modes does not necessary imply that perturbation at one site will propagate to the others. Many low-frequency modes must be considered to obtain proper convergence of the dynamics correlation.<sup>83</sup>

#### 3.5.4 Normal mode analysis (NMA)

The now classic method of normal mode analysis (NMA) is a powerful approach to analyze dynamical features of biological systems, particularly in characterizing large-scale conformational changes and granting valuable insight at timescales of microseconds or longer.<sup>20–22,31,32,84,85</sup> Normal modes approximate the motion of the system as harmonic vibrations around a local minimum on the potential energy surface. The essential tasks to calculate normal modes are first protein minimization, and second diagonalization of the second derivatives, often called the *Hessian*, to yield the normal modes and their frequencies which are the eigenvectors and eigenvalues, respectively.<sup>86</sup> The most difficulties arise from the diagonalization of the  $3N \times 3N$  *Hes-*

sian matrix, where  $N$  is the number of atoms. Storage of the *Hessian* requires large amounts of computer memory, and full-scale diagonalization is often not feasible for medium to large biomolecular systems. However, several methods have been put forward to reduce the size of calculation considerably without sacrificing the accuracy of the low-frequency modes, e.g. diagonalization in a mixed basis (DIMB)<sup>87,88</sup> and block normal mode (BNM)<sup>22,89</sup>. Further simplifications, e.g. the elastic network model (ENM)<sup>90</sup>, where the protein is modeled as a network of oscillators ( $C_\alpha$ -atoms) coupled together by uniform harmonic springs (within a given distance cutoff) are computationally inexpensive but still capture the system's intrinsic flexibility reasonably well. As long as the *Hessian* matrix is maintained, low-frequency modes are determined by the protein shape and are largely robust against coarse-graining levels and the potential function.<sup>23</sup> NMA provided useful insight into the nature of collective motions in proteins and its usefulness and limitations are described in a comprehensive review.<sup>24</sup>

### Simplification of complex protein motion

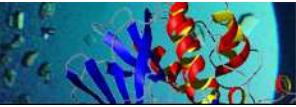
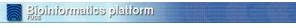



One of the main attractions of normal mode analysis is that the complex protein motion is simplified as a "superposition" of orthogonal, i.e. independent of each other, harmonic normal modes. Each normal mode (NM) coordinate specifies a set of atomic displacements at a specific frequency which can be analyzed separately. Low-frequency NMs are thermally well accessible and often correspond to functionally relevant motions. However, in no case does the lowest frequency NM correspond to such a functional mode and the challenge is to identify those involved in biological function.

*Involvement coefficient  $I_n$  for specific transition.* If two functional conformations are known, e.g. active and inactive state or liganded and unliganded structure, the contribution of each NM to this transition can be calculated.

$$I_n = \frac{\overrightarrow{X_1} - \overrightarrow{X_2}}{|\overrightarrow{X_1} - \overrightarrow{X_2}|} \cdot \overrightarrow{EV}_n \quad (3.6)$$

The involvement coefficient  $I_n$  quantifies the geometrical similarity between the  $n$ -th NM and the structural change, while  $\overrightarrow{X_1} - \overrightarrow{X_2}$  is the displacement vector between the two conformations and  $\overrightarrow{EV}_n$  the eigenvector of the  $n$ -th NM. Thus, one can sort the modes in terms of their contributions and concentrate on only those with the largest contributions. Statistically the motion lies most often along the direction of

**Table 2: Web-based programs to retrieve normal modes**

	<p><b>El Némo</b><sup>91</sup></p> <p><a href="http://igs-server.cnrs-mrs.fr/elnemo/start.html">http://igs-server.cnrs-mrs.fr/elnemo/start.html</a>            Elastic network model that provides several models based on low-frequency NMs together with their visualization.            Requires conformation, if two conformations are given the contribution to the transition is calculated additionally.</p>
	<p><b>WEBnm</b><sup>92</sup></p> <p><a href="http://services.cbu.uib.no/tools/normalmodes">http://services.cbu.uib.no/tools/normalmodes</a>            Requires conformation.</p>
	<p><b>NOMAD</b> (Normal Mode Analysis Deformation, and Definement)</p> <p><a href="http://lorentz.immstr.pasteur.fr/overlap/overlap_submission.php">http://lorentz.immstr.pasteur.fr/overlap/overlap_submission.php</a>            Uses elastic network model to calculate the overlap between the difference vectors of the two forms and each set of eigenvectors of lowest frequency normal modes.            Requires two structure superposed conformations.</p>
	<p><b>oGNM</b> (Normal Mode Analysis Deformation, and Definement)</p> <p><a href="http://ignm.ccbb.pitt.edu/GNM_Online_Calculation.htm">http://ignm.ccbb.pitt.edu/GNM_Online_Calculation.htm</a>            Requires conformation.</p>
	<p><b>AD-ENM Webserver</b></p> <p><a href="http://enm.lobos.nih.gov/">http://enm.lobos.nih.gov/</a>            Uses elastic network model.            Requires conformation.  <b>DC-ENM</b><sup>93</sup> - builds a structural model for a protein in an unknown state by using NMA and distance constraints.  <b>PATH-ENM</b><sup>94</sup> - generates transition path between two given structures.</p>

two modes. Though, to obtain few significant low-frequency modes the two conformations must differ substantially in structure which is the case for motor proteins, i.e. proteins that convert chemical to structural energy, but not in signaling proteins that undergo only subtle structural changes. Characterization of the NMs implicated in the conformational change give hints regarding hinge regions and collectively moving substructures.

#### NM frequency

The NM frequencies can be related to experimental spectroscopy techniques, such as infrared and Raman spectroscopy. Though one must be aware that the harmonic approximation may not provide an adequate physical model of a biological molecule under physiological conditions. In any case the frequency is an indicator of the entropy of the system. High frequencies describe a stiff system, whereas low-



frequencies are soft and involve many atoms being most relevant to allostery.

NMA provides motions that are easily accessible by the protein in a given conformation, i.e. around an energy minimum, although approximated by the use of harmonic functions. Even so, biologically interesting motions can be explored without the much more computationally expensive and time-consuming MD simulations. Definitely, NMA has provided a great deal of insight into the nature of collective motions in proteins.

### 3.5.5 Biased molecular dynamics simulations

Introducing a biased force into the conventional force field can significantly speed up processes that are not feasible yet with equilibrium MD simulations, e.g. ligand binding/unbinding processes.<sup>38,40,95</sup> The thereby sampled intermediate structures are events that presumably happen naturally on longer time scales and reveal details of molecular interactions such as sidechain rearrangements and H-bond networks. In addition, thermodynamic potentials can in principle be reconstructed even from irreversible processes.<sup>96,97</sup>

#### Steered molecular dynamics (SMD)

Steered MD simulations require "knowledge" of the direction in which the pulling force is exerted and is at best a straight line. Recording of ligand positions and applied forces reveal important structural information about binding pathways and interactions. In the avidin-biotin complex the ligand moves in discrete steps due to formation and rupture of a H-bond network.<sup>95</sup> It should be noted that the choice of the pulling force or speed must be chosen carefully and may be guided from experimental results, e.g. data from atomic force microscopy or optical tweezer experiments.

#### Targeted molecular dynamics (TMD)

Structural transition paths between two known conformations, an initial and a desired target structure, can be explored by the targeted MD method. To drive the system from one known state to another constraints of the form

$$V = k \cdot (x - x_{target})^2 \quad (3.7)$$

where  $k$  is the force constant and  $x$  and  $x_{target}$  are the atomic coordinates of the cur-

rent and target structure, are used. The protein's response to the induced change is analyzed. Biased simulations driving the repositioning of the  $\beta 4$ - $\alpha$  loop and the sidechain of residue Y106 were used to elucidate temporal relation and causality in the allosteric "Y-T coupling" mechanism in CheY.<sup>37</sup> It is essential to drive initial to end state transitions in both directions to ensure consistence of observed binding events. If a structural change is truly cooperative similar effects are expected if the induced and observed change are exchanged.

#### **Umbrella sampling to construct potential of mean force (PMF)**

Potential of mean force (PMF) is the energy profile along a predefined pathway. Along a meaningful reaction coordinate introduction of a harmonic potential ensures extensive sampling at a specific point, commonly termed umbrella sampling. Moving the position of the biased potential along the path while ensuring the windows are overlapping, data of interest is acquired rapidly, even sampling events that are otherwise rare. From these data free energy barriers of conformational transitions are calculated using the weighted histogram analysis method (WHAM). The energy barriers found can be compared to kinetics data and help to model the system's energy landscape.

#### **Anisotropic thermal diffusion (ATD) / pump-probe MD**

The mechanism of signal transduction in allostery is still heavily under investigation. The propagation of local perturbations may either be transmitted through coupled motion relying on global dynamics which are intrinsic to the protein shape and can be explored using NMA (see section 3.5.4), or the signal may follow a defined communication pathway. The latter is probed by visualizing the flow of excess energy from a specific site throughout the protein. The general idea is to introduce a perturbation at a specific residue or protein part, e.g. adding excess heat energy<sup>35</sup> or introducing a driving force<sup>36</sup>, and to monitor the propagation through the protein. Thermal diffusion can be observed if the protein is first equilibrated at low temperature, freezing atomic motion ( $T = 10$  K, RMSD  $< 0.05$  Å), and then coupling a selection to a high temperature bath ( $T = 300$  K).<sup>35</sup> The energy flow is monitored by conformational distortions from the starting structure after several time-delays, taking into consideration the elapsed time for a signal to travel through proteins. For good path resolution the use of implicit solvent, using a distance dependent dielectric, is necessary. The use of non-equilibrium simulations greatly enhances signal-to-noise and simultane-

ously decreases the simulation time needed ( $< 10$  ps). The found diffusion process is highly non-isotropic highlighting a residue network. Results for the PDZ domain family are in remarkable agreement with previous co-evolution studies<sup>33</sup>, but additionally indicated that van der Waals interactions is the main component of signal transduction. Thus fast and efficient intramolecular energy propagation relies on the precise positioning of residues and is sensitive to point mutations or minor structural changes such as sidechain orientations. Computational approaches are vital, providing a physical basis for signal propagation.

The main application of biased MD simulations is to drive the system through some specific transition or sequence of events which are not yet accessible to conventional MD. Using these techniques the protein's response to perturbation can be followed, and temporal relation and coupling of events in allosteric processes can be detected and even energetic information can be deduced. However, the remaining challenge is to find suitable reaction coordinates to obtain meaningful information. This problem is circumvented in the pump-probe approach where the flow of excess energy from a specific site throughout the protein directly highlights possible communication routes.

### 3.5.6 Domain identification

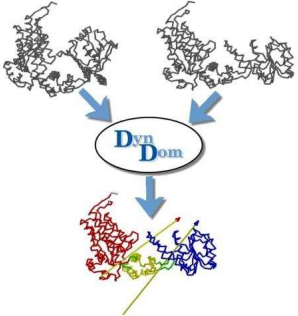
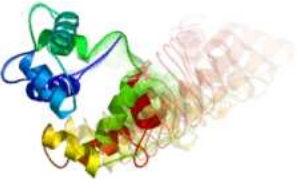


A key concept to study protein structure is the domain, a compactly folded region of a protein that has independent stability and is relatively rigid, which is jointed to other domains by flexible regions and can move relative to them. For example, the opening-closing mechanism to accommodate a substrate is a long-known functional motion emphasizing the structure-function relationship. Databases<sup>98-100</sup> store collections of potentially functional motions. Common protein motions identified are hinge bending resulting in an "open" and "closed" conformation, shear motion where one domain slides along the surface of another, and partial un-/refolding of the protein, the first two being illustrated in Figure 3 D.

Difference-distance plots and graph theoretic methods<sup>101</sup> are commonly used to identify rigid regions in proteins. Residues of rigid parts in the protein have small difference changes in distance between two conformations. Although easy to implement and fast, this method does not provide information about the geometric proximity of residues. In the graph theoretic method the protein is represented as a graph whose

### 3 Methods Survey

vertices are atoms and whose edges are distance constraints described by strong local forces. Rigidity theory determines the rigid clusters and the flexible joints connecting them, and has the advantage of requiring only a single conformation. In addition, web-based programs offer analysis and (animated) visualization for two conformations provided.

**Table 3: Web-based programs to highlight structural changes**

	<p><b>DynDom</b><sup>102,103</sup> <a href="http://fizz.cmp.uea.ac.uk/dyndom/">http://fizz.cmp.uea.ac.uk/dyndom/</a> Automatically determines domains, hinge axes and hinge bending residues in proteins. Requires two structural conformations. Output: domains and hinge residues colored, axis indicated by arrow.</p>
	<p><b>MolMovDB</b><sup>98,99</sup> (Database of Macromolecular Movements) <a href="http://www.molmovdb.org/">http://www.molmovdb.org/</a> Information on protein motion for given Protein Data Bank ID. Morph Server – generates and animates chemically realistic interpolations between two conformations. Hinge Prediction – predicts hinge locations in a single protein structure.</p>
	<p><b>DomainFinder</b><sup>104</sup> (Database of Macromolecular Movements) <a href="http://dirac.cnrs-orleans.fr/DomainFinder/">http://dirac.cnrs-orleans.fr/DomainFinder/</a> Variant of normal mode analysis based on deformation energy. Free download of program. Requires single structure (PDB format).</p>
	<p><b>DomainParser</b><sup>101</sup> (Database of Macromolecular Movements) <a href="http://compbio.ornl.gov/structure/domainparser/">http://compbio.ornl.gov/structure/domainparser/</a> Based on network flow problem. Requires Protein Data Bank ID.</p>

Characterization of conformational changes in terms of hinge axes, rotation points and shifting along surfaces offers a basis for mechanistic description of functional motion and critical residues for flexibility and interaction, e.g. hinge residues or H-bond networks, can be identified. The availability of web-based programs indicated in the list below, particularly animated movies describing a possible pathway between two structural conformations, has encouraged this approach to study protein allostery.

### 3.5.7 Calculation of energy contributions and hot-spot residue identification

The interactions between proteins and other molecules are critical to many processes in biology, and estimations of binding free energies are desirable. Approaches for estimating either relative or absolute binding free energies at different levels of accuracy and computational requirements have been developed. A widely applied technique is the molecular mechanics generalized Born surface area (MM-GBSA) model<sup>105</sup> which is computationally efficient and allows the decomposition of the binding free energy into residual contributions. This is especially useful to identify residues that contribute most to the association free energy<sup>105,106</sup>, so-called hot-spot residues<sup>107</sup>. In contrast to experimental mutagenesis experiments it is relatively straightforward to introduce mutations computationally, and hence a computational alanine scanning approach has been described.<sup>108,109</sup> Lately, methods were developed to estimate absolute binding free energies<sup>110–112</sup> that can directly be compared to experimentally measured binding constants, and the effect from displacing solvent molecules is being studied<sup>113</sup>. However, stringent accuracy is required to be of use in drug discovery where correct ranking of a list of ligands can be vital.

### 3.5.8 Bioinformatics: sequence-structure relationship

Knowledge of the three-dimensional structure of the protein is invaluable to start computational studies and to understand protein function in a mechanistic sense. Genome sequencing projects provided vast data of protein sequences. However, the majority of sequences remain without known 3D structure despite great advances in experimental structure determination such as X-ray crystallography, NMR and high-resolution electron microscopy. To bridge this gap a lot of effort has been put into the development of computational methods to predict 3D structures (see refs.<sup>114,115</sup> and the comprehensive book chapters<sup>116,117</sup>).

#### Homology modeling

Homology modeling, also known as comparative modeling or template-based modeling, attempts to construct atomic-resolution of a protein based on its amino acid sequence. The general procedure involves the selection of a known structure with



high sequence identity which is used as the template structure. Alignment of the template with the unknown so called target sequence is followed by the model construction. Coordinates are generated first for the structurally conserved core from those of the known structure. Structurally variable loops and side-chain conformations are built thereafter, and finally the model structure is refined. The success of homology modeling is based on the fact that evolutionary related sequences generally have similar 3D structures due to the constraints of proper protein folding and protein function. The final accuracy of the homology model is greatly affected by the sequence identity and the correctness of sequence alignment. If the sequence identity is below 30%, a region often referred to as the "twilight zone", serious errors such as misprediction of the protein fold can occur. If no suitable template structure is available for comparative modeling, *de novo* modeling methods also called *ab initio* modeling may be used. The quality of homology models does not allow processes such as drug-design, where the exact sidechain orientation is crucial. Nevertheless, the models are useful for qualitative conclusions and can guide mutagenesis experiments and structure-function hypotheses. In conjunction with molecular dynamics simulations, homology models are useful to obtain kinetics and dynamics information of the protein system.

The structural genomics project attempts to generate a diverse set of representative experimental structures for each class of protein folds, so that every newly discovered protein sequence is within modeling distance of a solved structure.<sup>118,119</sup> Thus, reliable structure prediction methods have become very important and are assessed in the bi-annual Critical Assessment of Techniques for Protein Structure Prediction (CASP) providing a community wide platform challenging prediction teams to predict structures which have been solved but not yet published. Because the volume of data is too large to process manually, automated structure prediction servers<sup>120</sup> are necessary.

#### **Folding of intrinsically disordered regions induced by target binding**

Intrinsically disordered proteins can adopt folded structures upon binding to their biological target. For example the TipAL protein, a bacterial transcriptional regulator, forms two additional  $\alpha$ -helices induced upon complex formation with the antibiotic, whereas in the apo structure this part is unstructured as revealed by NMR spectroscopy.<sup>121</sup> Based on the ligand-induced helix formation a possible mechanism for TipAL transcriptional activation was proposed where the introduced rigidity trans-





**Table 4: Web-based programs to generate homology models**

	<p><b>SwissModel</b><sup>120</sup></p> <p><a href="http://swissmodel.expasy.org/SWISS-MODEL.html">http://swissmodel.expasy.org/SWISS-MODEL.html</a>  A fully automated protein structure homology-modeling server.  Requires protein sequence or UniProt AC code.</p>
	<p><b>WHAT IF</b></p> <p><a href="http://swift.cmbi.kun.nl/servers/html/index.html">http://swift.cmbi.kun.nl/servers/html/index.html</a>  Protein structure analysis program for mutant prediction, structure verification and molecular graphics.  Requires aligned sequences and PDB file of template structure for homology modeling.</p>

mits forces to the DNA-binding parts. Intact secondary structure is also particularly important for immunological signal transduction. Molecular dynamics simulations showed that the small NC  $\beta$ -sheet is the key structural fragment to transmit conformational changes between domains V and C, the variable and constant domains of the F<sub>ab</sub> fragment of immunoglobulin G.<sup>82</sup> Coupled folding and binding is realized to be a general mechanism of molecular recognition<sup>122</sup>, where the inherent flexibility allows structural modification in response to different targets controlling cellular signaling. Attempts to predict regions of native disorder from sequence have been made.<sup>123–127</sup>

### Evolutionary conserved networks from sequence analysis

Evolution can be viewed as large-scale random mutagenesis with selection constraints imposed by protein function. Residues important for function, e.g. active site residues for enzymatic reactions or implicated in allosteric signalling, are not only highly conserved within domain families but also co-evolved. Ranganathan and co-workers<sup>33</sup> proposed a method to quantitatively calculate the level of conservation, comparing amino acid frequencies to their mean values after multiple sequence alignment, and to detect energetic coupling between two residues  $i$  and  $j$ . If sites  $i$  and  $j$  co-evolved, for functional or structural reason, a perturbation at site  $i$  will lead to a change in the amino acid distribution at site  $j$  highlighting their mutual dependence. This sequence-based statistical method was successfully applied to several protein families of distinct folds and functions<sup>33,34</sup> such as the protein domain PDZ, trans-

Table 5: Web-based programs to find potentially functional disordered parts	
	<p><b>DISOPRED2</b><sup>123</sup></p> <p><a href="http://bioinf.cs.ucl.ac.uk/disopred/disopred.html">http://bioinf.cs.ucl.ac.uk/disopred/disopred.html</a>            Identifies dynamically disordered regions with no stable secondary structure and high flexibility in solution.            Requires protein sequence.</p>
	<p><b>DisEMBL</b><sup>124</sup></p> <p><a href="http://dis.embl.de">http://dis.embl.de</a>            Predicts intrinsic protein disorder.            Requires protein sequence or SwissProt ID.</p>
	<p><b>DisProt</b></p> <p><a href="http://www.disprot.org">http://www.disprot.org</a>            Database of protein disorder providing information about proteins that lack fixed 3D structure in their putatively native states, either in their entirety or in part.            Requires search entry by keyword or sequence.</p>
	<p><b>RONN</b><sup>125</sup> (Regional Order Neural Network)</p> <p><a href="http://www.strubi.ox.ac.uk/RONN">http://www.strubi.ox.ac.uk/RONN</a>            Database of protein disorder providing information about proteins that lack fixed 3D structure in their putatively native states, either in their entirety or in part.            Requires protein sequence.</p>
	<p><b>GLOBPLOT2</b><sup>126</sup> (Intrinsic Protein Disorder, Domain &amp; Globularity Prediction)</p> <p><a href="http://globplot.embl.de">http://globplot.embl.de</a>            Identifies the tendency of order/globularity and disorder.            Requires sequence or SwissProt ID.</p>
	<p><b>Domain prediction from sequence</b></p> <p><b>DomCut</b><sup>128</sup></p> <p><a href="http://www.bork.embl.de/~suyama/domcut/">http://www.bork.embl.de/~suyama/domcut/</a>            Identifies domain linkers.            Requires protein sequence.</p> <p><b>DomPred</b><sup>128</sup></p> <p><a href="http://bioinf.cs.ucl.ac.uk/dompred/DomPredform.html">http://bioinf.cs.ucl.ac.uk/dompred/DomPredform.html</a>            Identifies domain linkers.            Requires sequence.</p>

membrane signaling receptors GPCRs or the multi-subunit protein hemoglobin. This method can be automated using cluster analysis of perturbation matrices for protein families to map the global network of amino acid interactions. This method allows identification of potentially important residues from sequence only. Nevertheless,



knowledge of structure contributes strongly to the visualization, identified residues that often form pathways of connected residues, and helps to find a signal transduction mechanism.

Profound understanding of the relationship between protein sequence and structure is of major interest. In particular bioinformatics tools attempt to close the gap between the rapid protein sequencing and the slow structure determination. Additionally, the protein sequence can tell us about intrinsically flexible regions that are potentially functional and long-range communication routes that have developed during evolution.

#### **Contributions from computational methods:**

- bioinformatics assists in providing structural models and bridges the gap between vast protein sequence data and very limited protein structures
- MD gives structure at atomistic level and dynamics simultaneously
- probe signaling mechanisms: energy flow or coupled motion
- provide energetics information: binding affinity and mapping the protein energy landscape
- dissect complex motion into simple movements of individual NMs and assist in the mechanistic description of domain motion



# 4 Bacterial Second Messenger c-di-GMP and PleD as Model System

## 4.1 Ubiquitous bacterial second messenger c-di-GMP

Cyclic di-guanosine-monophosphate (c-di-GMP) was first discovered as an activator of cellulose synthase in *Acetobacter xylinum* (currently *Gluconacetobacter xylinus*) in the 1980's by Benziman and co-workers.<sup>129</sup> However, its impact as a major cell regulator was discovered only recently.<sup>130–133</sup> C-di-GMP is a novel second messenger in bacteria, but not in eukaryotes or Archae and intense research is conducted to elucidate its biological function and the interplay with cellular components<sup>134,135</sup>. To date, c-di-GMP is known to orchestrate cellular pathways implicated in bacterial cell adhesion and motility, including the swarmer-to-stalk cell transformation<sup>136,137</sup>, biofilm formation<sup>138,139</sup>, pathogenicity and virulence<sup>140</sup> described in recent reviews<sup>141,142</sup> and references therein.

### **DGCs, PDEs and c-di-GMP control the concentration of the second messenger**

High cellular levels of c-di-GMP inhibit motility and activate biofilm formation, a surface-attached bacterial community embedded in an extracellular matrix. Thus, to secure adequate regulation of the transition between motile and sessile lifestyles a tight control of the cellular c-di-GMP concentration is critical. The c-di-GMP concentration is regulated by the opposing activity of diguanylate cyclases (DGCs), which synthesize c-di-GMP from two GTP molecules, and c-di-GMP specific phosphodiesterases (PDEs) cleaving the cyclic compound into the inactive linear pGpG (see Figure 5 A).<sup>143</sup> Genetic analysis and biochemical characterization demonstrated that DGC activity resides exclusively in the GGDEF domain (earlier termed DUF1)<sup>131,144–149</sup>, named after the conserved sequence motif that constitutes part of the active site. PDE

## 4 Bacterial Second Messenger c-di-GMP and PleD as Model System

---

activity is assigned to the EAL domain (earlier termed DUF2).<sup>144,150–152</sup> The high abundance of GGDEF domains in the bacterial kingdom, often associated with the EAL domain, found by subsequent genome screens underlines the importance of the second messenger c-di-GMP and its regulation by these domains. In addition, c-di-GMP itself can modulate DGC activity by allosteric control. The ability for product inhibition was first discovered in the multidomain protein PleD of *Caulobacter crescentus*<sup>51</sup> and later identified as a general feature, intrinsic to the GGDEF domain and requiring the motif RxxD of conserved residues in the inhibition site (see section 5.1).<sup>148</sup> Based on the X-ray structure of PleD with bound c-di-GMP a "domain immobilization" model was proposed, where c-di-GMP links domains preventing the productive encounter of two loaded active sites (illustrated in Figure 6 B).<sup>51,52</sup>

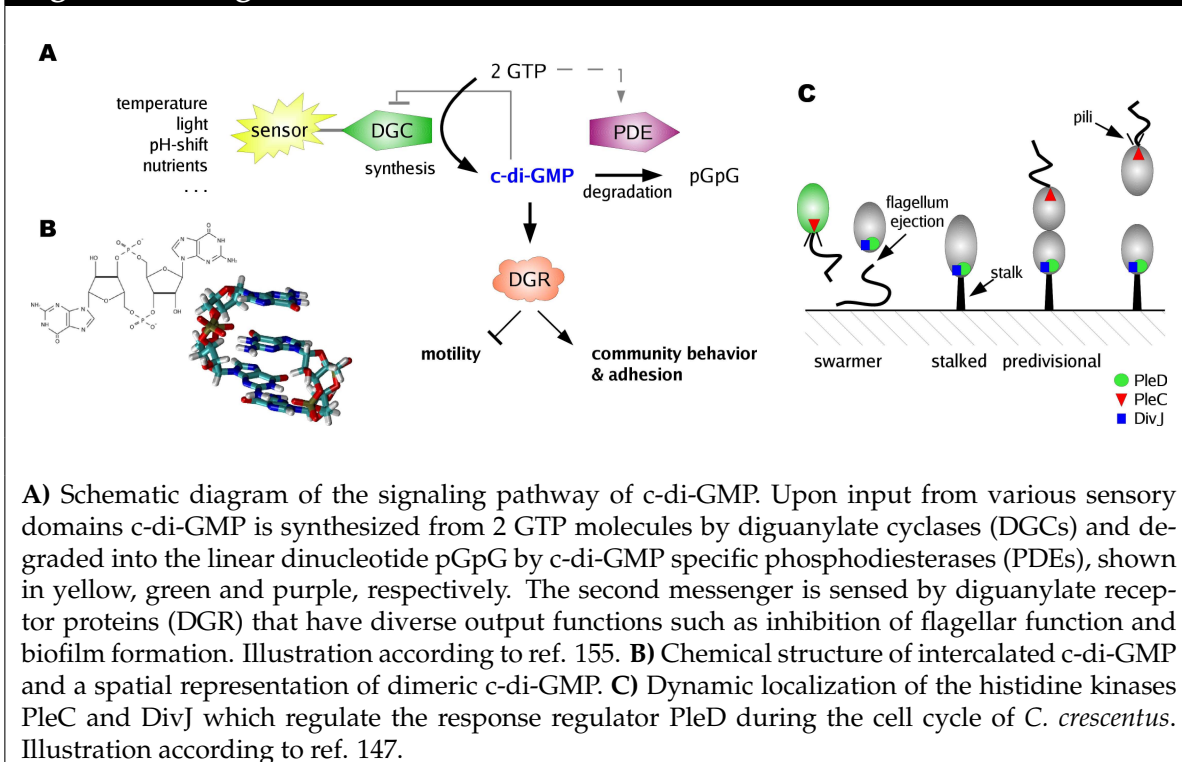
### C-di-GMP network

Fusion of the catalytically active domain GGDEF to numerous sensory domains<sup>147</sup> allows the conversion of environmental stimuli such as change in temperature, light, pH-shift, availability of oxygen and nutrients into internal signals via synthesis of small diffusible molecules, here c-di-GMP. The variety of sensory domains combined with GGDEF domains creates a complex network of regulatory interactions and may link different signaling pathways. The search for c-di-GMP downstream elements has just begun, and the PilZ protein family was recently identified as a class of specific diguanylate receptors.<sup>153–156</sup>

The demonstration that binding of c-di-GMP to an I-site of the GGDEF domain prevents further production of c-di-GMP is of medicinal relevance and renders DGCs a potential target to combat biofilm-related infections. However, unlike the established c-di-GMP controlled switch from motile single-cell lifestyle to surface-attached multicellular communities, a complete picture of the regulatory pathways of c-di-GMP and a detailed understanding of the mechanism of allosteric control of c-di-GMP on the GGDEF domain are still lacking. In the following PleD of *C. crescentus*, serving as a model system for which extensive structural and biochemical data is available, is characterized in detail and computational results on the allosteric regulation are presented in section 5.

**Biological function of the second messenger:**

- c-di-GMP regulates motility, biofilm formation and pathogenicity
- c-di-GMP concentration is controlled by DGC and PDE activity
- c-di-GMP inhibits DGC activity which is intrinsic to the GGDEF domain
- c-di-GMP binds to the PilZ domain, a downstream receptor family

**Figure 5: Biological function of c-di-GMP**

## 4.2 PleD activity under allosteric control

### Diguanylate cyclase PleD is a response regulator

The multidomain protein PleD of *Caulobacter crescentus* is a well characterized diguanylate cyclase.<sup>51,52,136,147,148,157,158</sup> The domain architecture and a structural representation of the protein is given in Figure 6 A and D. The C-terminal catalytically active GGDEF domain (DGC) synthesizes the second messenger c-di-GMP as output signal and in turn regulates the pole remodeling which is required for the swarmer-to-stalk

## 4 Bacterial Second Messenger c-di-GMP and PleD as Model System

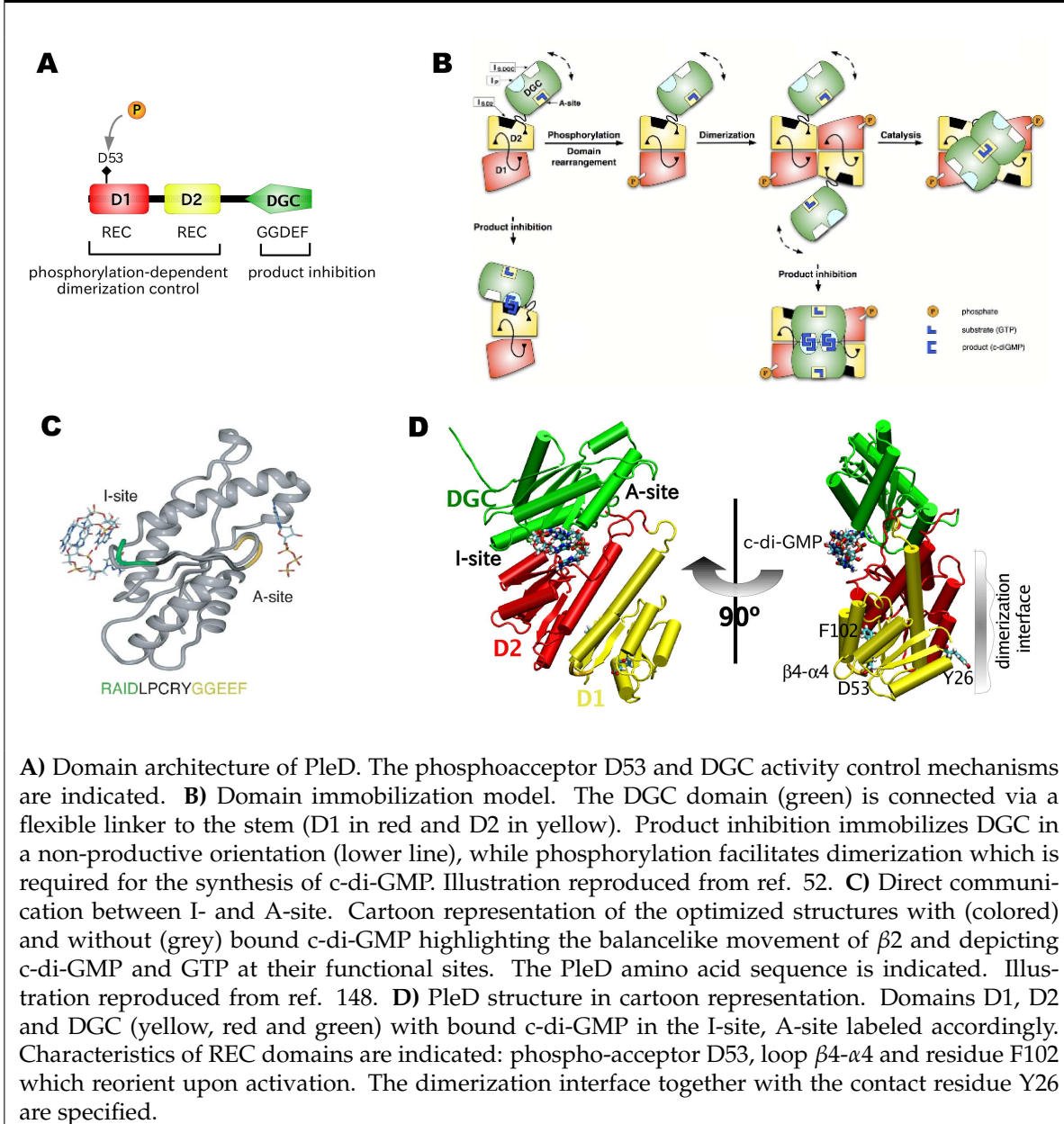
---

cell transition in the bacterial lifestyle (see Figure 5 C).<sup>136,137,159,160</sup> The DGC domain is fused to two N-terminal CheY-like receiver domains arranged in tandem (D1 and D2), which have the  $(\beta\alpha)_5$  topology typical for response regulator receiver domains (REC).<sup>51,136</sup> The domain architecture of N-terminal sensory domains linked to a C-terminal effector domain is characteristic for response regulator proteins. Response regulators (RR) are part of a phosphotransfer pathway, commonly referred to as two-component signal transduction, which is widespread in bacteria to "process" external stimuli.<sup>161,162</sup> Changes in the environment are sensed by kinases, usually located in the membrane, and trigger the transfer of a phosphate group to a conserved residue on the RR. Phosphorylation induces a conformational change in the regulatory domain that results in the activation of an associated domain that effects the response, and is the most common covalent modification used to achieve allosteric control in proteins. PleD is phosphorylated and dephosphorylated at residue D53 of the first REC domain D1 by the action of the DivJ and PleC histidine kinases, respectively. Fluorescence experiments showed co-localization of DivJ with PleD to the flagellated pole restricting c-di-GMP production spatially and triggering flagella ejection and stalk synthesis.<sup>147,157,163</sup>

### Inhibition via domain immobilization

Structure determination of non-phosphorylated PleD (code 1W25<sup>51</sup>), co-crystallized with c-di-GMP identified an allosteric site for the mutually intercalated c-di-GMP dimer, here referred to as c-di-GMP though in reality it is  $[\text{c-di-GMP}]_2$ . Binding of c-di-GMP to this allosteric site (inhibition site, I-site) leads to strong non-competitive inhibition of PleD as demonstrated by kinetic studies.<sup>51,148</sup> C-di-GMP binds to the DGC domain through residues R359, D362 and R390. Moreover, c-di-GMP present in the I-site can crosslink to either the neighboring domain D2 or to another close-by DGC domain by additional interactions to residues R148 and R178 or R313, respectively.<sup>51,52</sup> Based on these structural findings a mechanistic model was suggested where bound c-di-GMP locks the DGC domain in a non-productive orientation which blocks the encounter of two GTP-loaded active sites and subsequent conversion to c-di-GMP (see Figure 6 B). The two different immobilization modes found, namely DGC-D2 and DGC-DGC crosslinking, are redundant and the integrity of only one is required for non-competitive product inhibition. The occurrence of DGC-DGC immobilization extends the applicability of the proposed mechanistic model of domain immobilization even to single domain proteins.

Figure 6: Regulation of PleD



**A)** Domain architecture of PleD. The phosphoacceptor D53 and DGC activity control mechanisms are indicated. **B)** Domain immobilization model. The DGC domain (green) is connected via a flexible linker to the stem (D1 in red and D2 in yellow). Product inhibition immobilizes DGC in a non-productive orientation (lower line), while phosphorylation facilitates dimerization which is required for the synthesis of c-di-GMP. Illustration reproduced from ref. 52. **C)** Direct communication between I- and A-site. Cartoon representation of the optimized structures with (colored) and without (grey) bound c-di-GMP highlighting the balancelike movement of  $\beta 2$  and depicting c-di-GMP and GTP at their functional sites. The PleD amino acid sequence is indicated. Illustration reproduced from ref. 148. **D)** PleD structure in cartoon representation. Domains D1, D2 and DGC (yellow, red and green) with bound c-di-GMP in the I-site, A-site labeled accordingly. Characteristics of REC domains are indicated: phospho-acceptor D53, loop  $\beta 4$ - $\alpha 4$  and residue F102 which reorient upon activation. The dimerization interface together with the contact residue Y26 are specified.

### Product inhibition is intrinsic to the GGDEF domain

Subsequent studies on the single domain diguanylate cyclase DgcA of *C. crescentus* confirmed unambiguously that product inhibition is a general feature of DGCs and is intrinsic to the GGDEF domain.<sup>148</sup> Thus, feedback inhibition allows a tight control of the cellular level of c-di-GMP defining an upper limit for the concentration of the second messenger. The nature of communication between I- and A-site necessary for

self-regulation is still being investigated. The results presented in this work argue for a combination of structural rearrangement and simultaneous changes in protein flexibility. Comparison of optimized structures of PleD with and without the ligand c-di-GMP present in the I-site showed a balancelike movement of the  $\beta$ -strand  $\beta 2$  which connects the two functional sites and may in turn affect the catalytic activity by displacing the enzymatic residues (see Figure 6 C).<sup>148</sup> Moreover, a conserved RxxD sequence motif in the I-site ( $\alpha 2$ - $\beta 2$  loop) together with a conserved spacer length between the I- and A-site is found in more than 60% of GGDEF proteins. In addition to the structural aspect, dynamical coupling between the two sites is revealed by normal mode analysis, i.e. upon c-di-GMP binding to the allosteric site a marked drop in protein motion is observed for both I- and A-site simultaneously.<sup>81</sup> Binding of c-di-GMP also has long-range effects changing the domain-domain correlations, DGC-D2 coupling is enhanced where D1-D2 become less correlated. Further research on the allosteric control in PleD and GGDEF domains in general is required. In particular, the exact role of the RxxD motif remains to be investigated.

### **Additional level of control via phosphorylation-mediated dimerization**

Consistent with the symmetric condensation reaction to synthesize c-di-GMP, cross-linking experiments demonstrated that DGC activity of PleD is contained entirely within the dimer fraction.<sup>158</sup> The dimerization rate and hence the c-di-GMP production depends on the phosphorylation-state of PleD and it appears that this regulation is the sole function of the first REC domain D1 and, while D53N mutation of the phosphoryl acceptor site abolishes activation ability completely. Unlike domain D1 which contains all functional residues of a RR<sup>50</sup> (D9, D10, phosphoacceptor D53, T83, F102 and K105), the role of the second REC domain is unclear and it was proposed to function as an adaptor.<sup>51,157</sup>

In enzymatic studies and structure determinations the unstable phosphate group is often replaced with  $\text{BeF}_3^-$ , which is a stable functional and structural analog of the phosphoryl group.<sup>50,164-167</sup> Indeed,  $\text{BeF}_3^-$ -activated PleD is approximately 35 times more active than unphosphorylated PleD<sup>158</sup>, and could be crystallized successfully<sup>52</sup>. In terms of structure, activation leads to a D1/D2 domain repositioning eventually promoting tight homodimerization as indicated by a 500  $\text{\AA}^2$  increase of the buried surface per monomer. Structural changes upon activation of receiver domains have been studied thoroughly on CheY<sup>50,168,169</sup> and other receiver domains<sup>161,170</sup> detecting the characteristic  $\beta 4$ - $\alpha 4$  loop repositioning accompanied by the sidechain rotation of



a conserved Tyr/Phe residue (F102 in PleD) and a concomitant repositioning of Thr (T83) to interact with the phospho moiety. F102 being located in the D1/D2 interface potentially influences the domain rearrangement facilitating dimerization. However, little is known about how these well known conformational changes influence dimerization and are transmitted to the effector domain. In addition, the functional role of residue Y26 which is required for PleD dimerization is completely unknown.<sup>158</sup>

Combined genetic, biochemical and structural data showed two independent layers of DGC control, product inhibition and phosphorylation-mediated dimerization. These allosteric regulations of DGC activity raise the question of how the information is transferred between the distant functional sites. PleD X-ray structures<sup>51,52</sup> being available opens the possibility to complement experimental data with computer simulations providing dynamics information together with the structure at atomic resolution. In the following sections, the effect of bound c-di-GMP is studied by normal mode analysis and cross-correlation maps derived from trajectories. Furthermore, the signal transmission in the dimerization process is investigated on a truncated PleD system, considering simply the D1/D2 stem domains. Due to the substantial orientational freedom of the DGC domain *vs.* the stem domains that form the dimer interface, the DGC domain is neglected. Here, we focus on events following the local  $\beta$ 4- $\alpha$ 4 loop reorientation with regard to PleD dimerization.

### Allosteric regulation in PleD

- "domain immobilization": c-di-GMP locks domains in non-productive orientation
- phosphorylation-mediated dimerization
- dynamic coupling between inhibition- and active-site (this work)
- change in interdomain correlated motion (this work)
- extended  $\alpha$ 5-helix has leverage effect (this work)



# 5 Results

## 5.1 Allosteric control of cyclic di-GMP signaling

B. Christen, M. Christen, R. Paul, F. Schmid, M. Folcher, P. Jenoe,  
M. Meuwly and U. Jenal  
**J. Biol. Chem.** (2006) 281:32015-32024

My contribution: Computational simulations and molecular modeling.

Allosteric Control of Cyclic di-GMP Signaling<sup>\*[5]</sup>

Received for publication, April 13, 2006, and in revised form, June 21, 2006. Published, JBC Papers in Press, August 21, 2006, DOI 10.1074/jbc.M603589200

Beat Christen<sup>†1</sup>, Matthias Christen<sup>†1</sup>, Ralf Paul<sup>‡</sup>, Franziska Schmid<sup>§</sup>, Marc Folcher<sup>‡</sup>, Paul Jenoe<sup>‡</sup>, Markus Meuwly<sup>§</sup>, and Urs Jenal<sup>†2</sup>From the <sup>†</sup>Biozentrum and the <sup>§</sup>Department of Chemistry, University of Basel, Klingelbergstrasse 70, 4056 Basel, Switzerland

Cyclic di-guanosine monophosphate is a bacterial second messenger that has been implicated in biofilm formation, antibiotic resistance, and persistence of pathogenic bacteria in their animal host. Although the enzymes responsible for the regulation of cellular levels of c-di-GMP, diguanylate cyclases (DGC) and phosphodiesterases, have been identified recently, little information is available on the molecular mechanisms involved in controlling the activity of these key enzymes or on the specific interactions of c-di-GMP with effector proteins. By using a combination of genetic, biochemical, and modeling techniques we demonstrate that an allosteric binding site for c-di-GMP (I-site) is responsible for non-competitive product inhibition of DGCs. The I-site was mapped in both multi- and single domain DGC proteins and is fully contained within the GGDEF domain itself. *In vivo* selection experiments and kinetic analysis of the evolved I-site mutants led to the definition of an RXXD motif as the core c-di-GMP binding site. Based on these results and based on the observation that the I-site is conserved in a majority of known and potential DGC proteins, we propose that product inhibition of DGCs is of fundamental importance for c-di-GMP signaling and cellular homeostasis. The definition of the I-site binding pocket provides an entry point into unraveling the molecular mechanisms of ligand-protein interactions involved in c-di-GMP signaling and makes DGCs a valuable target for drug design to develop new strategies against biofilm-related diseases.

A global signaling network that relies on the production of the second messenger cyclic diguanylic acid has recently been discovered in bacteria (1, 2). The c-di-GMP<sup>3</sup> system emerges as a regulatory mastermind orchestrating multicellular behavior and biofilm formation in a wide variety of bacteria (2). In addition, c-di-GMP signaling also plays a role in bacterial virulence

and persistence (3–7). The broad importance of this novel signaling molecule in pathogenic and non-pathogenic bacteria calls for careful analysis of the molecular mechanisms that control cellular levels of c-di-GMP and regulate its downstream targets. c-di-GMP is formed by the condensation of two GTP molecules (8–10) and is hydrolyzed to GMP via the linear intermediate pGpG (11–14). Two widespread and highly conserved bacterial protein domains have been implicated in the synthesis and hydrolysis of c-di-GMP, respectively (15). The breakdown of c-di-GMP is catalyzed by the EAL domain (12–14), and the diguanylate cyclase (8) activity resides in the GGDEF domain (10, 16). The highly conserved amino acid sequence GG(D/E)EF forms part of the catalytically active site (A-site) of the DGC enzyme (8). In agreement with this, mutations that change the GG(D/E)EF motif generally abolish the activity of the respective proteins (14, 16–18).

GGDEF domains are often found associated with sensor domains, arguing that DGC activity is controlled by direct signal input through these domains (1). The best understood example for controlled activation of a DGC is the response regulator PleD, which constitutes a timing device for *Caulobacter crescentus* pole development (17, 19, 20). PleD is activated during *C. crescentus* development by phosphorylation of an N-terminal receiver domain and, as a result, sequesters to the differentiating cell pole (17, 19). An additional layer of control was suggested by the crystal structure of PleD solved recently in complex with c-di-GMP (8) (Fig. 1). A c-di-GMP binding site was identified in the crystal, spatially separated from the catalytically active site (A-site). Two mutually intercalating c-di-GMP molecules were found tightly bound to this site, at the interface between the GGDEF and the central receiver-like domain of PleD (Fig. 1). Based on the observation that PleD activity shows a strong non-competitive product inhibition, it was proposed that this site might constitute an allosteric binding site (I-site) (8). Based on the observation that functionally important residues of the PleD I-site are highly conserved in a majority of GGDEF proteins listed in the data base, we tested the hypothesis that allosteric product inhibition is a general regulatory principle of bacterial diguanylate cyclases.

## MATERIALS AND METHODS

*Strains, Plasmids, and Media*—*Escherichia coli* and *Salmonella enterica* serovar Typhimurium strains were grown in Luria broth (LB). *C. crescentus* strains were grown in complex peptone yeast extract (21). For DGC activity assays *in vivo*, *E. coli* was plated onto LB Congo Red plates (Sigma, 50 µg/ml). To determine the IPTG induction phenotype, 3 µl of a liquid log phase culture was spotted onto LB Congo Red plates with-

\* This work was supported by Swiss National Science Foundation Fellowship 3100A0-108186 (to U. J.) and by a Swiss National Science Foundation Förderprofessur (to M. M.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[5] The on-line version of this article (available at <http://www.jbc.org>) contains supplemental text and Figs. S1–S4.

<sup>1</sup> These authors contributed equally to this work.

<sup>2</sup> To whom correspondence should be addressed: Tel: 41-61-267-2135; Fax: 41-61-267-2118; E-mail: urs.jenal@unibas.ch.

<sup>3</sup> The abbreviations used are: c-di-GMP, cyclic diguanylic acid; pGpG, linear diguanylic acid; LB, Luria broth; DGC, diguanylate cyclase; PDE, phosphodiesterase; H6, hexa-histidine tag; rdar, red, dry, and rough; IPTG, isopropyl 1-thio-β-D-galactopyranoside; DgcA, diguanylate cyclase A; PdeA, phosphodiesterase A; CR, Congo Red; AC, adenylate cyclase; GC, guanylate cyclase.

### Diguanylate Cyclase Feedback Control

out and with 1 mM IPTG. Biofilm formation was quantified after overnight growth by staining with 1% Crystal Violet as described (22). Motility phenotypes were determined using LB or peptone yeast extract motility plates containing 0.3% Difco-Agar. The exact procedure of strain and plasmid construction is available on request.

**Random I-site Tetrapeptide Library**—The *dgca* gene (CC3285) was amplified by PCR using primers #1006 and #1007 (for primer list see supplemental text). The PCR product was digested with NdeI and XhoI and cloned into pET21a (Novagen). In a next step a *dgca*ΔRESΔ allele with a silent PstI restriction site was generated by splicing with overlapping extension PCR using primers #1129, #670, and #1132. The resulting PCR product was digested with NdeI and XhoI and cloned into pET42b (Novagen) to produce pET42::*dgca*ΔRESΔ. The PstI/XhoI fragment of pET42b::*dgca*ΔRESΔ was replaced by 20 independent PCR products, which had been generated using pET42b::*dgca*ΔRESΔ as a template and primers #1131 and #670. The resulting 20 independent random libraries were individually transformed into *E. coli* BL21 and screened on Congo Red plates (LB plates supplemented with 50 μg/ml Congo Red). As a control reaction, the deleted I-site was reverted back to the wild-type RESΔ motif by cloning the PCR product generated with primers #1130 and #670 into the PstI and XhoI site of pET42b::*dgca*ΔRESΔ.

**Diguanylate Cyclase and Phosphodiesterase Activity Assays**—DGC reactions were performed at 30 °C with 0.5 μM purified hexahistidine-tagged DgcA or 5 μM PleD in DGC reaction buffer containing 250 mM NaCl, 25 mM Tris-Cl, pH 8.0, 5 mM β-mercaptoethanol, and 20 mM MgCl<sub>2</sub>. For inhibition assays the protein was preincubated with different concentrations of *c*-di-GMP (1–100 μM) for 2 min at 30 °C before 100 μM [<sup>33</sup>P]GTP (Amersham Biosciences) was added. The reaction was stopped at regular time intervals by adding an equal volume of 0.5 M EDTA, pH 8.0. DGC/PDE tandem assays were carried out using 1 μM hexahistidine-tagged DgcA, which was preincubated for 2 min in the presence or absence of 4.5 μM hexahistidine-tagged phosphodiesterase PdeA. The reaction was started by adding 100 μM [<sup>33</sup>P]GTP. The reactions were stopped at regular time intervals of 15 s by adding equal volumes of 0.5 M EDTA, pH 8.0, and their nucleotide composition was analyzed as described below.

Initial velocity ( $V_0$ ) and inhibition constants were determined by plotting the corresponding nucleotide concentration *versus* time and by fitting the curve according to allosteric product inhibited Michaelis-Menten kinetics with the program ProFit 5.6.7 (with fit function  $[c\text{-di-GMP}]_t = a(1)^*t/(a(2) + t)$ , where the initial velocity  $V_0$  is defined as  $a(1)/a(2)$ ) using the Levenberg-Marquardt algorithm.  $K_i$  values were determined by plotting  $V_0$  *versus* *c*-di-GMP concentration and using the following fit function,  $V_{0[c\text{-di-GMP}]} = V_{0[c\text{-di-GMP}] = 0} * (1 - ([c\text{-di-GMP}]/(K_i + [c\text{-di-GMP}])))$ .

**Polyethyleneimine Cellulose Chromatography**—Samples were dissolved in 5 μl of running buffer containing 1:1.5 (v/v) saturated NH<sub>4</sub>SO<sub>4</sub> and 1.5 M KH<sub>2</sub>PO<sub>4</sub>, pH 3.60, and blotted on Polygram® CEL 300 polyethyleneimine cellulose TLC plates (Macherey-Nagel). Plates were developed in 1:1.5 (v/v) saturated NH<sub>4</sub>SO<sub>4</sub> and 1.5 M KH<sub>2</sub>PO<sub>4</sub>, pH 3.60 ( $R_f(c\text{-di-GMP})$  0.2,

$R_f(pGpG)$  0.4), dried, and exposed on a storage phosphor imaging screen (Amersham Biosciences). The intensity of the various radioactive species was calculated by quantifying the intensities of the relevant spots using ImageJ software version 1.33.  $V_0$  and  $K_i$  were determined with the Software ProFit 5.6.7.

**UV Cross-linking with [<sup>33</sup>P]*c*-di-GMP**—The <sup>33</sup>P-labeled *c*-di-GMP was produced enzymatically using [<sup>33</sup>P]GTP (3000 Ci/mmol) and purified according to a previous study (14). Protein samples were incubated for 10 min on ice in DGC reaction buffer (25 mM Tris-HCl, pH 8.0, 250 mM NaCl, 10 mM MgCl<sub>2</sub>, 5 mM β-mercaptoethanol) together with 1 μM *c*-di-GMP and <sup>33</sup>P-radiolabeled *c*-di-GMP (0.75 μCi, 6000 Ci/mmol). Samples were then irradiated at 254 nm for 20 min in an ice-cooled, parafilm-wrapped 96-well aluminum block in an RPR-100 photochemical reactor with a UV lamp RPR-3500 (Southern New England Ultraviolet Co.). After irradiation, samples were mixed with 2× SDS-PAGE sample buffer (250 mM Tris-HCl at pH 6.8, 40% glycerol, 8% SDS, 2.4 M β-mercaptoethanol, 0.06% bromophenol blue, 40 mM EDTA) and heated for 5 min at 95 °C. Labeled proteins were separated by SDS-PAGE and quantified by autoradiography.

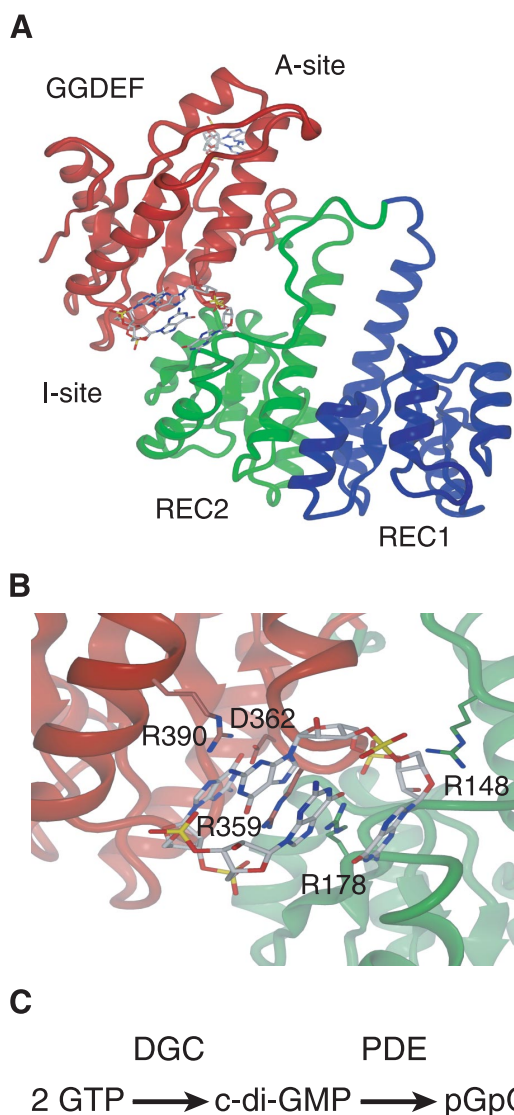
**Nucleotide Extraction and Analysis**—2.0 ml of *E. coli* cell cultures ( $A_{600}$  0.4) were harvested by centrifugation, and supernatant was discarded. The cell pellet was dissolved in 200 μl of 0.5 M formic acid, and nucleotides were extracted for 10 min at 4 °C. Insoluble cell components were then pelleted, and the supernatant was directly analyzed by chromatography. Nucleotides were extracted and separated according to a previous study (23) on a 125/4 Nucleosil 4000-1 polyethyleneimine column (Macherey-Nagel) using the SMART-System (Amersham Biosciences). The nucleotide peak corresponding to *c*-di-GMP was verified by co-elution with a chemically synthesized *c*-di-GMP standard.

**DgcA Protein Expression Levels**—DgcA protein expression levels in *E. coli* BL21 were determined by Western blot analysis using Anti-His(C-Term) antibody (Invitrogen) and horseradish peroxidase conjugate of goat anti-mouse IgG (Invitrogen) as secondary antibody. The protein concentration was determined by measuring the intensities of the relevant spots using ImageJ software version 1.33. Signals were calibrated to defined concentrations of purified wild-type DgcA.

**Molecular Modeling of PleD**—All-atom simulations were carried out using the CHARMM (24) program and the CHARMM22/27 force field (25). For additional information see the supplemental material.

## RESULTS

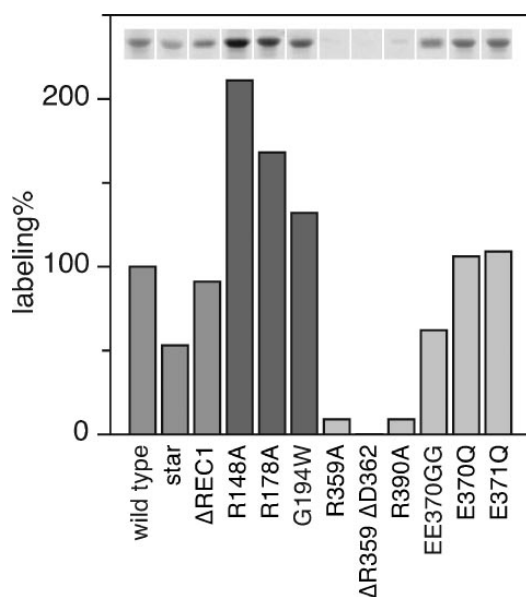
**Feedback Inhibition of the PleD Diguanylate Cyclase Requires Binding of *c*-di-GMP to the I-site**—The PleD crystal structure indicated the existence of an allosteric binding pocket (I-site) at the interface of the GGDEF and REC2 domains (8). Binding of a *c*-di-GMP dimer in the I-site is mediated by specific electrostatic interactions with charged residues of the GGDEF and REC2 domain (Fig. 1). To provide evidence for *c*-di-GMP binding to the I-site pocket in solution, trypsin digests were performed with purified PleD protein (5 μM) in the presence or absence of *c*-di-GMP (25 μM). The resulting peptide fragments were separated on a C18 column and analyzed by matrix-as-



**FIGURE 1. Crystal structure of the response regulator PleD.** A, domain architecture of PleD with receiver domain REC1 (blue), receiver domain REC2 (green), and GGDEF domain harboring diguanylate cyclase activity (red). The active site (A-site) loop and the allosteric binding site (I-site) are indicated. B, zoom in view of the I-site pocket with a bound dimer of c-di-GMP with intercalated purine bases. Residues Arg-148 and Arg-178 (green) from the REC2 domain and residues Arg-359, Asp-362, and Arg-390 (red) from the GGDEF domain make specific contacts to the ligand in the crystal structure. C, schematic of c-di-GMP synthesis and degradation reactions.

sisted laser desorption ionization time-of-flight. Both chromatograms were identical, with the exception of two peaks that were only detected in the absence of ligand but were protected when c-di-GMP present during tryptic digest (supplemental Fig. S1). One of the two peptides (T47, retention time 25.6 min) was identified by mass spectrometry and corresponds to the amino acids 354–359 (supplemental Fig. S1), arguing that c-di-GMP specifically protects from trypsin cleavage at Arg-359. To

### Diguanylate Cyclase Feedback Control



**FIGURE 2. c-di-GMP labeling efficiency of different PleD mutants.** The upper lane shows autoradiographs of [<sup>33</sup>P]c-di-GMP UV cross-linked hexahistidine-tagged PleD mutant proteins separated by SDS-PAGE. Relative labeling efficiency with c-di-GMP is shown below with wild-type PleD corresponding to 100%. Specific mutants in different domains are colored in gray (REC1), dark gray (REC2) and light gray (GGDEF).

provide additional evidence for ligand binding in solution, we performed UV cross-linking assays using <sup>33</sup>P-labeled c-di-GMP (14). Residues Arg-148 and Arg-178 of the REC2 domain and Arg-359, Asp-362, and Arg-390 of the GGDEF domain were replaced with alanine, and the resulting protein variants were analyzed. As shown in Fig. 2, mutating I-site residues of the GGDEF domain abolished (ΔR359ΔD362) or strongly reduced (R359A and R390A) c-di-GMP binding. In contrast, mutations in the A-site (E370Q, E371Q, and EE370GG), which completely abolished enzymatic activity (Table 1), had no effect on c-di-GMP binding (Fig. 2), indicating that labeling with radioactive c-di-GMP results from ligand binding at the I-site. Although mutations R359A, R359V, ΔR359ΔD362, and D362A all showed a dramatically reduced or complete loss of enzymatic activity, mutant R390A showed wild-type-like DGC activity (Table 1). In agreement with the reduced binding of c-di-GMP (Fig. 2), the *K<sub>i</sub>* of mutant R390A was increased ~20-fold (Table 1). PleD proteins harboring mutations in the REC2 portion of the I-site (R148A and R178A) showed an increased binding of c-di-GMP (Fig. 2) and slightly lower *K<sub>i</sub>* values than wild type (Table 1). Surprisingly, R148A/R178A single and double mutants displayed a 5- to 20-fold higher DGC activity compared with wild-type PleD (Table 1). Finally, c-di-GMP binding was normal in mutant proteins that either lacked the REC1 receiver domain or had a bulky tryptophan residue introduced at the REC2-GGDEF interface (G194W, Fig. 2). Together these results implied that the structural requirements for c-di-GMP binding are contained within the GGDEF domain of PleD and that residues Arg-359, Asp-362, and Arg-390 form the core of an allosteric binding pocket for c-di-GMP.

## Diguanilate Cyclase Feedback Control

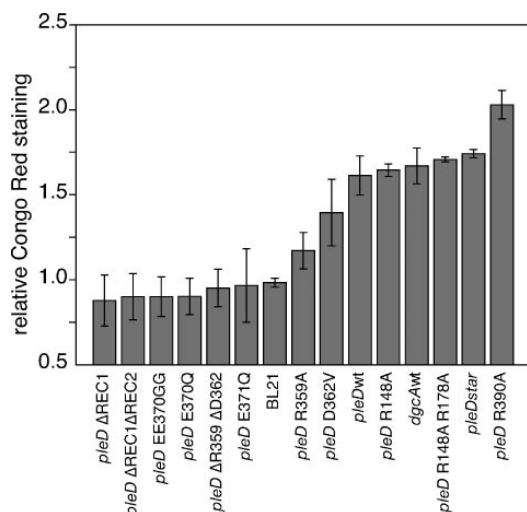


FIGURE 3. *In vivo* activity of different PleD and DgcA mutant proteins. *E. coli* BL21 strains expressing different *pleD* alleles and *dgcA* wild type were spotted onto Congo Red plates. Relative Congo Red binding was determined using imageJ software with BL21 corresponding to 100%.

*Evidence for an in Vivo Role of I-site-mediated Feedback Control*—To test a possible role for feedback inhibition of diguanilate cyclases *in vivo*, we developed a simple assay based on the observation that in *E. coli* and other *Enterobacteriaceae* increased cellular levels of c-di-GMP correlate with Congo Red (CR) staining of colonies on plates (28). Low level expression (in the absence of the inducer IPTG) of active *pleD* alleles caused a red colony phenotype in the *E. coli* B strain BL21, whereas cells expressing inactive *pleD* alleles under the same conditions stained white (Fig. 3). Interestingly, PleD mutants with dramatically different diguanilate cyclase activities *in vitro* showed only minor differences of CR staining *in vivo*. For instance, PleDR148A/R178A, which showed a 20-fold increased activity (Table 1), or PleD\*, a constitutively active mutant of PleD several 100-fold more active than wild-type (9), caused virtually identical CR values like PleD wild type (Fig. 3). In contrast, expression of the feedback inhibition mutant PleDR390A resulted in a significantly higher CR staining even though its *in vitro* DGC activity was lower than wild-type PleD (Table 1). This argued that *in vivo* steady-state concentrations of c-di-GMP were determined mainly by the PleD inhibition constant (as opposed to the overall activity of the enzyme) and that a functional I-site is critical for DGC control *in vivo*.

*DgcA, a Single Domain Diguanilate Cyclase, Is Subject to Allosteric Product Inhibition*—Sequence alignments of >1000 annotated GGDEF domain proteins revealed that that I-site residues Arg-359 and Asp-362 of PleD are highly conserved. 57% of the proteins containing a GGDEF domain and 27% of GGDEF/EAL composite proteins possess this motif. This suggested that c-di-GMP product inhibition could be a general regulatory mechanism of bacterial diguanilate cyclases. To test this, hexahistidine-tagged derivatives of two *C. crescentus* GGDEF domain proteins were analyzed biochemically with respect to their DGC activities and c-di-GMP binding proper-

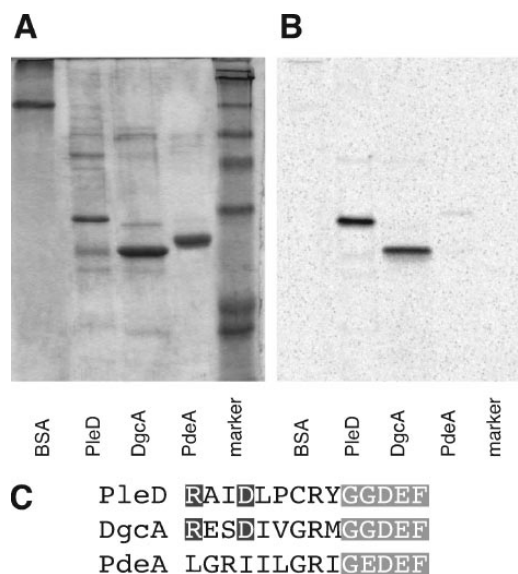


FIGURE 4. UV cross-linking of different GGDEF domains with  $^{33}\text{P}$ -labeled c-di-GMP. A, Coomassie-stained SDS-PAGE and B, autoradiograph of BSA (control), PleDΔREC1, DgcA, and the isolated GGDEF domain of the c-di-GMP-specific phosphodiesterase PdeA (CC3396) after UV cross-linking with  $^{33}\text{P}$ -c-di-GMP. C, alignment of I- and A-site sequence of PleD, DgcA, and PdeA. I-site (RXXD) and A-site residues (GGDEF) are marked in black and gray, respectively.

ties. Purified DgcA (diguanilate cyclase A, CC3285), a soluble, single domain GGDEF protein that lacks an obvious N-terminal input domain, showed strong diguanilate cyclase activity (Fig. 5A). DgcA has an RESD motive five amino acids upstream of the conserved GGDEF active site and was readily labeled with  $^{33}\text{P}$ -c-di-GMP in a cross-linking experiment (Fig. 4). Consistent with this, DgcA showed strong feedback inhibition (Fig. 5A) with its  $K_i$  (1  $\mu\text{M}$ ) being in the same range as the inhibition constant determined for PleD (8). In contrast, the GGDEF domain of PdeA (phosphodiesterase A, CC3396), which lacks catalytic activity (14), had no conserved I-site residues and did not bind radiolabeled c-di-GMP (Fig. 4). Thus, specific binding of c-di-GMP correlated with the presence of a conserved I-site motif RXXD (Fig. 4).

Diguanilate cyclase activity assays revealed strong and rapid product inhibition of DgcA. DgcA alone was able to convert only a small fraction of the available GTP substrate pool into the product c-di-GMP ( $V_o = 2.8 \mu\text{mol}$  of c-di-GMP  $\mu\text{mol}$  protein $^{-1}$  min $^{-1}$ ) (Fig. 5B). In contrast, GTP consumption and conversion into c-di-GMP and pGpG was rapid ( $V_o = 43.0 \mu\text{mol}$  of c-di-GMP  $\mu\text{mol}$  protein $^{-1}$  min $^{-1}$ ) and almost complete when the PDE CC3396 was added in excess to the enzymatic reaction (Fig. 5B). This argued that c-di-GMP feedback inhibition is abolished in a sequential DGC-PDE reaction, because the steady-state concentration of the inhibitor c-di-GMP is kept low by continuous degradation of c-di-GMP into the linear dinucleotide pGpG. As a consequence of rapid feedback inhibition, the experimentally determined  $V_o$  values of the DGC reaction are generally underestimated. In conclusion, these results strengthen the view that allosteric product inhibi-

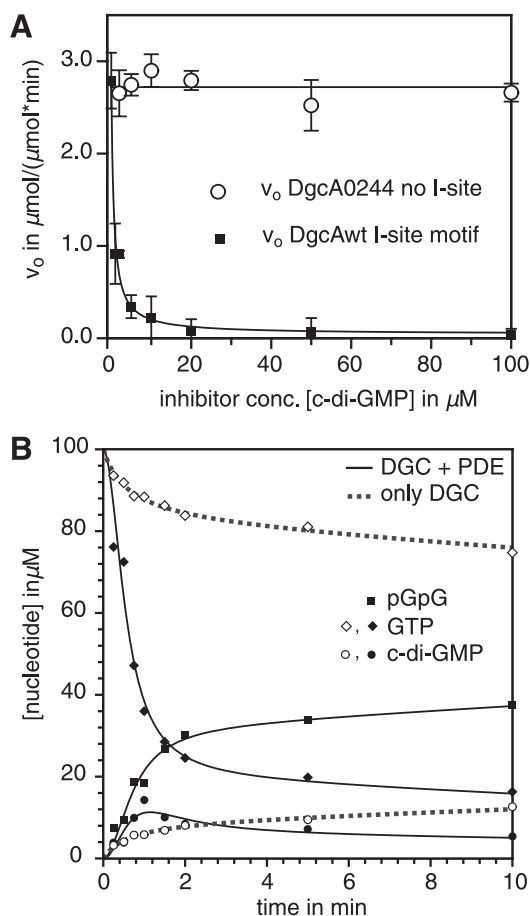


FIGURE 5. **c-di-GMP product inhibition of DgcA.** A, initial velocities of the wild-type diguanylate cyclase DgcA (squares) and the non-feedback inhibited I-site mutant DgcA0244 (circles) in the presence of increasing concentrations of c-di-GMP. B, conversion of GTP into c-di-GMP by DgcA (dashed lines) and accelerated GTP consumption, c-di-GMP synthesis, and cleavage into pGpG by a diguanylate cyclase-phosphodiesterase tandem reaction (plain lines).

tion is a general principle of diguanylate cyclases and that high affinity binding of c-di-GMP requires an RXXD I-site motif positioned in close proximity to the active site.

**Development of an *in Vivo* Assay to Genetically Probe Allosteric Control of DgcA**—DGCs from different bacterial species have been shown to be functionally interchangeable (17, 26, 27). To determine if DgcA is active *in vivo* we expressed a plasmid-based copy of the *dgcA* gene in *C. crescentus*, *S. enterica*, and *Escherichia coli* B and tested the respective strains for the phenotypes known to result from increased cellular levels of c-di-GMP (17, 26, 27). Consistent with these earlier findings, expression of *dgcA* strongly inhibited flagellar-based motility in all three organisms, dramatically increased the ability of *S. enterica* and *E. coli* for surface colonization, and produced the characteristic red, dry, and rough (rdar) colony morphology when plated on CR plates (Fig. 6, A–F) (29). The red phenotype provided the basis for a visual genetic screen on CR plates. Under these conditions, cells producing active DgcA variants would

produce dark red single colonies, whereas cells producing inactive DgcA mutants would remain white. This prompted us to use the CR screen to isolate *dgcA* mutants, which had a specific defect in feedback regulation, and to define the minimal requirements for product inhibition of this class of enzymes.

**Randomization of c-di-GMP Binding Pocket Reveals Three Mutant Classes**—To probe the minimal requirements of the I-site for c-di-GMP binding and product inhibition, a *dgcA* mutant library was constructed with the RESD signature replaced by a randomized tetrapeptide sequence (see “Materials and Methods”). In short, a *dgcA* gene, which carried a deletion of the four I-site codons, was used as template for a PCR reaction. For the amplification step a primer complementary to the 3′-end of *dgcA* was used in combination with a mixture of oligonucleotides that spanned the deletion site and contained 12 randomized base pairs at the position coding for the deleted amino acids. The resulting PCR fragments were fused in-frame with the 5′-end of *dgcA* in the expression plasmid pET42b and were transformed into *E. coli* BL21. The resulting gene library contains a theoretical number of  $1.67 \times 10^7$  ( $4^{12}$ ) different *dgcA* alleles, coding for DgcA variants with different combinations of I-site residues.

When plated on CR plates, colonies transformed with a wild-type *dgcA* allele showed the typical rdar colony morphology (Fig. 6G). Transformation of *E. coli* BL21 with a plasmid expressing a mutant DgcA, which lacked the four amino acids of the I-site (DgcA $\Delta$ RESD), produced white colonies on CR plates (Fig. 6H), indicating that this mutant form had lost DGC activity. About 10% of the clones with random tetrapeptide insertions stained red on CR plates and thus had retained DGC activity (Fig. 6I). This result is consistent with the observation that alanine scanning of the PleD I-site almost exclusively produced non-active enzyme variants (Table 1) and argues that the majority of amino acid substitutions introduced at the I-site are detrimental for the catalytic activity of the DGC. To further characterize active DgcA I-site variants, a total of 800 red colonies was isolated and patched onto CR plates without (Fig. 6, J and K) or with the inducer IPTG (Fig. 6, L and M). This secondary screen was based on the observation that IPTG-induced expression of the *pleDR390A* allele (Table 1), but not of the *pleD* wild-type allele, abolished growth of *E. coli* BL21 (data not shown). This suggested that at elevated protein levels, DGCs that lack feedback control are toxic *in vivo* (see below). The majority of the I-site library clones tested failed to grow on plates containing IPTG, indicating that their activity is no longer controlled by product inhibition (Fig. 6, L and M). Only 7 mutants (out of 9000 colonies screened) showed a wild type-like behavior in that they stained dark red on CR plates and tolerated the presence of the inducer IPTG (Fig. 6, L and M).

This genetic screen led to the isolation of three different I-site mutant classes with the following characteristics: 1) catalytically inactive mutants ( $A^-$ , frequency  $\sim 90\%$ ); 2) feedback control negative mutants ( $I^-A^+$ , frequency  $\sim 10\%$ ); and wild-type-like mutants ( $I^+A^+$ , frequency  $\sim 0.1\%$ ). A subset of class 1 and 2 mutants and all seven class 3 mutants were selected, and hexahistidine-tagged forms of the respective proteins were purified for biochemical characterization. Kinetic parameters of activity ( $V_0$ ) and feedback inhibition ( $K_i$ ) were determined



## Diguanilate Cyclase Feedback Control

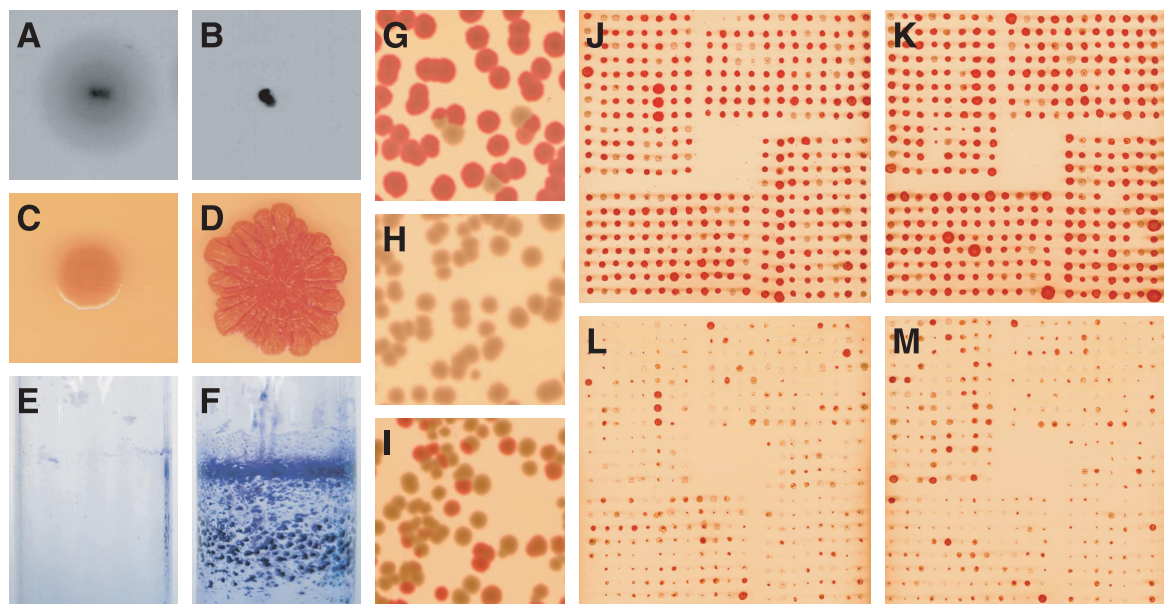


FIGURE 6. Phenotypic characterization of ectopically expressed diguanilate cyclase *dgca* in *E. coli* and *S. enterica*. Behavior of *E. coli* strain BL21 with empty pET42b plasmid (A) and pET42b::*dgca* (B) on motility plates. Colony morphology of *E. coli* strain BL21 with empty pET42b plasmid (C) and with pET42b::*dgca* (D) on Congo Red plates. Biofilm formation of *S. enterica* serovar Typhimurium *trp::T7RNAP* with empty pET42b (E) and pET42b::*dgca* (F) grown in liquid culture and stained with crystal violet. *E. coli* BL21 transformed with PCR-restored *dgca* wild type on pET42b::*dgca* (G), with the inactive allele *dgca*ΔRESΔ (pET42b::*dgca*ΔRESΔ) (H), and with a library of random tetrapeptide insertions in the I-site (pET42b::*dgca*XXXX) (I) and plated on Congo Red plates. *E. coli* BL21 expressing different I-site mutant alleles were spotted onto Congo Red plates without (J and K) and with 1 mM IPTG (L and M) to screen for feedback inhibition *dgca* alleles.

TABLE 1  
Kinetic analysis of PleD mutants

Protein	$V_o$	$\Delta V_o$	$K_i$	$\Delta K_i$
	$\mu\text{mol } c\text{-di-GMP/}$ $(\mu\text{mol protein} \cdot \text{min})$		$\mu\text{M}$	
PleD wild type	0.202	$\pm 0.023$	5.8	$\pm 1.0$
PleDR359A	0.005	ND <sup>a</sup>	>100	ND
PleDR359V	0.0	ND		
PleDΔ359Δ362	0.0	ND		
PleDD362A	0.0	ND		
PleDR390A	0.076	$\pm 0.007$	115.0	$\pm 18.1$
PleDR148A	0.822	$\pm 0.020$	2.8	$\pm 1.2$
PleDR178A	0.918	$\pm 0.292$	3.6	$\pm 0.1$
PleDR148AR178A	3.75	$\pm 0.43$	2.9	$\pm 0.6$
PleDG194W	0.161	$\pm 0.005$	6.3	$\pm 1.9$
PleDEE370GG	0.0	ND		
PleDE370Q	0.0	ND		
PleDE371Q	0.0	ND		

<sup>a</sup> ND, not determined.

individually using an *in vitro* diguanilate cyclase activity assay (16). Consistent with their *rdar*-like *in vivo* phenotype, only class 2 and class 3 mutants showed detectable diguanilate cyclase activity with an initial velocity between 1.93 and 14.21  $\mu\text{mol}$  of *c*-di-GMP  $\mu\text{mol protein}^{-1} \text{min}^{-1}$  (Table 2). Only mutant proteins from the IPTG tolerant class 3 showed product inhibition with  $K_i$  values close to 1  $\mu\text{M}$  (Table 2). In contrast, all proteins from class 2 mutants showed no feedback inhibition *in vitro*, arguing that their *in vivo* toxicity is the result of uncontrolled run-off DGC activity (Fig. 5A and Table 2). Support for this hypothesis comes from experiments determining the cellular concentration of *c*-di-GMP and DgcA protein expression levels in *E. coli* BL21 carrying selected *dgca* alleles on plasmid pET42b (see "Materials and Methods"). Alleles *dgca*A0244,

*dgca*I229, and *dgca*I250 were chosen, because the DGC activity of these enzymes is similar to wild type DgcA (Table 2). Basal level expression (no IPTG) of *dgca*A0244, the allele coding for a DGC that completely lacks feedback inhibition, resulted in a more than 100-fold increased cellular level of *c*-di-GMP as compared with cells expressing wild-type *dgca* (Table 3). This was due to an almost 100-fold higher overall turnover of the mutant enzyme as compared with wild type (Table 3). In contrast, enzymatic turnover and cellular concentration of *c*-di-GMP was increased only marginally in *E. coli* cells expressing alleles *dgca*I229, and *dgca*I250 with restored feedback inhibition control (Table 3).

Sequence analysis of the tetrapeptide insertions of class 2 and class 3 mutants revealed several important characteristics of a functional allosteric I-site binding pocket. All catalytically active and feedback inhibition competent mutants restored the wild-type Arg and Asp residues at positions one and four of the RXXD motive (Table 2). Whereas most of the mutants that had lost feedback inhibition had altered either one or both of these charged residues (Table 2) only two feedback inhibition mutants had retained both charges with changes in the intervening residues (Table 2). Obviously, Arg and Asp, while being strictly required for feedback inhibition, need to be placed in the appropriate sequence context of the I-site loop. These experiments define the minimal requirements of the I-site core region and demonstrate that the Arg and Asp residues that make direct contacts to the *c*-di-GMP ligand in the crystal structure are of critical functional importance for DGC feedback inhibition *in vivo* and *in vitro*. This provides a plausible

**TABLE 2**  
Diguanylate cyclase activity and inhibition constant of DgcA I-site mutant proteins

Protein	Motif	$V_o$ $\mu\text{mol } c\text{-di-GMP}/$ $(\mu\text{mol protein}\cdot\text{min})$	$\Delta V_o$	$K_i$ $\mu\text{M}$	$\Delta K_i$
DgcA wt	RESD	2.79	$\pm 0.01$	0.96	$\pm 0.09$
DgcA1406	RQGD	5.35	$\pm 0.05$	7.02	$\pm 2.92$
DgcA1040	RLVD	4.92	$\pm 0.19$	4.52	$\pm 1.81$
DgcA1229	RGAD	2.03	$\pm 0.01$	1.84	$\pm 0.26$
DgcA1524	RSAD	3.70	$\pm 0.13$	7.36	$\pm 2.69$
DgcA1529	RLAD	2.79	$\pm 0.04$	1.01	$\pm 0.23$
DgcA0751	RCAD	3.65	$\pm 0.10$	3.51	$\pm 0.52$
DgcA1250	RGGD	2.07	$\pm 0.02$	2.24	$\pm 0.49$
DgcA $\Delta$ RESD		0.14	$\pm 0.06$		ND <sup>a</sup>
DgcA0207	GMGG	14.21	$\pm 0.54$	No inhibition	
DgcA0244	VMGG	2.57	$\pm 0.05$	No inhibition	
DgcA0613	GGVA	4.29	$\pm 0.06$	No inhibition	
DgcA0646	GRDC	8.90	$\pm 0.10$	No inhibition	
DgcA0913	GVGD	3.81	$\pm 0.04$	No inhibition	
DgcA1300	MEGD	0.87	$\pm 0.02$	No inhibition	
DgcA1733	GMNH	11.47	$\pm 0.17$	No inhibition	
DgcA3018	RESE	11.1	$\pm 0.11$	No inhibition	
DgcA0230	RNRD	3.02	$\pm 0.06$	No inhibition	
DgcA0642	RVDS	4.17	$\pm 0.08$	No inhibition	
DgcA1007	RAGG	6.06	$\pm 0.05$	No inhibition	
DgcA2006	RGQD	1.93	$\pm 0.01$	No inhibition	

<sup>a</sup> ND, not determined.**TABLE 3**  
DgcA protein levels and cellular *c*-di-GMP concentrations in the absence or presence of IPTG induction at 1 mM

	Protein conc. <sup>a</sup>		<i>c</i> -di-GMP conc.		Turnover <sup>b</sup>	
	No induction	1 mM IPTG	No induction	1 mM IPTG	No induction	1 mM IPTG
	<i>pmol protein/mg dry weight</i>		<i>pmol c-di-GMP/mg dry weight</i>		<i>pmol c-di-GMP per pmol protein</i>	
DgcA0244	4.1	22	1466.3	1570.7	357.6	71.4
DgcA1229	3.5	31	87.6	139.5	25.0	4.5
DgcA1250	2.7	43	24.2	305.4	9.0	7.1
DgcA wt	2.9	33	13.75	189.4	4.7	5.7
DgcA $\Delta$ RESD	3.5	23	ND <sup>c</sup>	ND	NA <sup>d</sup>	NA

<sup>a</sup> See "Materials and Methods."<sup>b</sup> As derived from the cellular *c*-di-GMP concentration divided by the cellular protein concentration.<sup>c</sup> ND, not detectable.<sup>d</sup> NA, not applicable.

explanation for the strong conservation of the RXXD motif in GGDEF domains.

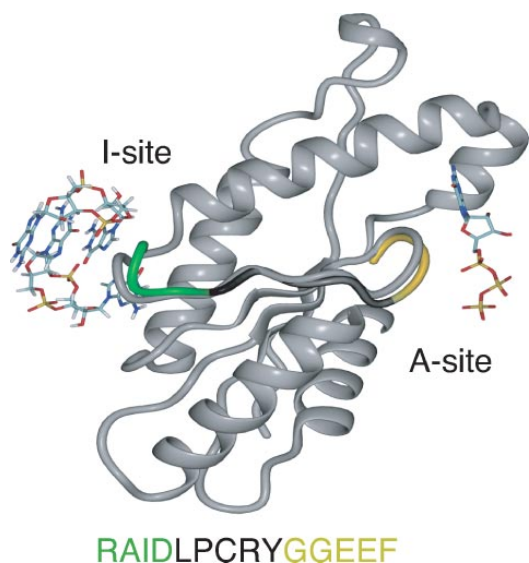
The molecular mechanism of product inhibition through I-site binding remains unclear. To assist the interpretation of the present data and provide information on binding induced mobility, atomistically detailed simulations were carried out. Normal mode calculations on ligated and unligated PleD were used to analyze the structural transitions that occur during I-site binding of *c*-di-GMP. Normal mode calculations on the optimized structures yielded no imaginary frequencies, and translational and rotational frequencies were close to zero ( $|\omega| \leq 0.02 \text{ cm}^{-1}$ ). This indicated that the minimized structures correspond to real minima on the potential energy surface. The displacements calculated for the ligated and the unligated protein showed a significant decrease in mobility for both I- and A-site residues upon complexation (supplemental Figs. S2 and S3). Whereas motion in the I-site is suppressed due to steric interactions upon ligand insertion, quenching of the A-site residues suggested that the two sites might be dynamically coupled via the short connecting  $\beta$ -strand ( $\beta_2$ ). Backbone  $C\alpha$ -atoms and side chains of the I-site and A-site loops were displaced by an average of 1–4 Å in opposite directions, arguing that a balance-like movement centered around  $\beta_2$  could be responsible for direct information transfer between the two sites (Fig. 7). The cumulated displacements per residue over all 147 modes

(supplemental Fig. S3) showed different mobilities in additional regions of the protein. The  $C\alpha$  atoms of residues exhibiting large changes in flexibility upon ligand binding are depicted as *spheres* in supplemental Fig. S3. Reduced flexibility (*yellow spheres*) is found at the I-site, A-site, phosphorylation site, and the dimer interface, whereas the flexibility is enhanced (*black spheres*) at the REC1/REC2 interface. In summary, these simulations show that I-site binding of *c*-di-GMP not only reduced the mobility around the RXXD motif but also of the residues of the A-site loop.

## DISCUSSION

*Feedback Inhibition Is a General Control Mechanism of Diguanylate Cyclases*—The data presented here propose a general mechanism to regulate the activity of diguanylate cyclases (DGCs), key enzymes of *c*-di-GMP-based signal transduction in bacteria. High affinity binding of *c*-di-GMP to a site distant from the catalytic pocket (I-site) efficiently blocks enzymatic activity in a non-competitive manner. Mutational analysis of multi- and single-domain DGC proteins has provided convincing evidence for the role of several charged amino acids in *c*-di-GMP binding and allosteric regulation. Furthermore, these experiments indicated that the allosteric binding site is functionally contained within the GGDEF domain. An *in vivo* selection experiment using a random tetrapeptide library, and

## Diguanylate Cyclase Feedback Control



**FIGURE 7. Comparison of the energy-minimized structures of the PleD GGDEF domain with and without ligand bound to the I-site.** For improved clarity, the domain is sliced through the I-site loop/ $\beta$ 2/A-site loop plane. The unligated protein is shown in gray and the I-site loop (green),  $\beta$ 2 (black), and A-site loop (gold) of the bound structure are shown as an overlay. GTP bound to the active site is modeled according to the orientation of c-di-GMP bound to the A-site in the crystal structure. The PleD amino acid sequence of I-site,  $\beta$ 2, and A-site is indicated below.

designed to re-engineer the I-site has led to the definition of a highly conserved RXXD core motif of the c-di-GMP binding pocket. The RXXD motif forms a turn at the end of a short five-amino acid  $\beta$ -sheet that directly connects the I-site with the conserved catalytic A-site motif, GG(D/E)EF (Fig. 7). This raised the question of how I-site ligand binding modulates DGC enzyme activity. In the multidomain protein PleD, c-di-GMP bound to the I-site physically connects the GGDEF domain with the REC1-REC2 dimerization stem. It was speculated that product inhibition occurs by domain immobilization, which would prevent the encounter of the two DGC substrate binding sites (8). Several observations argue in favor of a more direct communication between I- and A-sites. First, with a large variety of domains found to be associated with GGDEF domains, it seems unlikely that functional I-sites are generally formed by the interface of a GGDEF with its neighboring domain (2). In agreement with this, residues of the PleD REC2 domain are not required for c-di-GMP binding and feedback inhibition. Second, the single domain DGC protein, DgcA, shows I-site-dependent allosteric control with a  $K_i$  of 1  $\mu$ M. Third, the introduction of a bulky tryptophan residue (G194W) at the GGDEF-REC2 interface did not affect activity, I-site binding, or feedback inhibition of PleD (Fig. 2 and Table 1). Fourth, atomistic simulations of ligated and unligated PleD predicted a marked drop in flexibility of C $\alpha$ -atoms both in the I- and A-site upon ligand binding. Simultaneous with motion quenching,  $\beta$ 2 and its flanking I- and A-loops undergo a balance-like movement that repositions A-site residues in the catalytic active site (Fig. 7). This is consistent with the idea that structural changes within the GGDEF domain upon binding of c-di-GMP at the

I-site lead to repositioning of active site residues and possibly altered kinetic parameters. Thus, we propose that c-di-GMP binding and allosteric control represents an intrinsic regulatory property of DGCs that contain an RXXD motif.

Like guanylate and adenylate cyclases (GCs and ACs) and DNA polymerases, DGCs catalyze the nucleophilic attack of the 3'-hydroxyl group on the  $\alpha$ -phosphate of a nucleoside triphosphate. Despite the lack of obvious sequence similarities, the PleD x-ray structure revealed that DGCs possess a similar domain architecture like ACs and GCs (8, 30). Based on mutational analysis (8, 14, 16) and on structural comparisons between DGC, AC, GC, and DNA polymerases (31–34), a model for DGC catalysis can be proposed. In contrast to the heterodimeric ACs and GCs, DGCs form homodimers, with a GTP molecule bound within the catalytic core of each DGC monomer (8). Two Mg<sup>2+</sup> ions are coordinated by the highly conserved glutamic acid residue Glu-371, which is part of the GGDEF motif, and possibly by Asp-327 on the opposing  $\beta$ -sheet. The divalent Mg<sup>2+</sup> carboxyl complex coordinates the triphosphate moiety of GTP and activates the 3'-hydroxyl group for intermolecular nucleophilic attack. Substrate specificity of AC and GC can be interchanged by converting a few key residues involved in purine recognition (31, 34, 35). This includes an arginine residue, which in PleD corresponds to the highly conserved Arg-366 located in the  $\beta$ -sheet connecting the I- and A-sites. Based on the active site model, two alternative inhibition mechanisms can be envisaged. In a first scenario, binding of c-di-GMP to the I-site would change the orientation of Arg-366 and would thereby disturb the guanine binding pocket resulting in an increased  $K_m$  for GTP. Alternatively, inhibitor binding could rearrange the Mg<sup>2+</sup> carboxyl complex and thus destabilize the active state.

*In Silico Analysis of the GGDEF Protein Family Indicates That Product Inhibition Is a General Regulatory Mechanism*—DGC activity of GGDEF domain proteins seems to strictly depend on conserved GGDEF or GGEEF motifs in the active site (10, 16, 18, 36–38). Consistent with this, ~90% of the GGDEF and 62% of the GGDEF/EAL composite proteins show a conserved GG(D/E)EF A-site motif. Of the GGDEF proteins with a highly conserved A-site motif, >60% have conserved RXXD I-site residues and a conserved spacer length between I- and A-site, arguing that the three-dimensional arrangement of catalytic and allosteric pocket is likely to be similar in all DGCs. From a total of 19 GGDEF proteins, for which convincing evidence for a DGC activity exists, 14 have a conserved I-site (supplemental Fig. S4). Ryjenkov and coworkers (10) reported severe toxicity problems when expressing diguanylate cyclases lacking I-site residues in *E. coli* BL21. This is consistent with the growth defect observed upon expression of *dgcA* feedback inhibition mutants in *E. coli* BL21 and argues that these proteins are not feedback-controlled. The molecular basis of growth interference under these conditions is unclear. It is possible that depletion of the GTP pool or adverse effects of unphysiologically high levels of c-di-GMP are responsible for this effect. Although the experiments presented here define a role for the I-site in DGC feedback inhibition, the c-di-GMP binding pocket could also be exploited for other roles in c-di-GMP signaling. It has been proposed recently that non-catalytic GGDEF

domains with variant A-site motifs can fulfill regulatory functions (14). It is attractive to speculate that a subgroup of GGDEF proteins that has degenerate catalytic A-sites but conserved c-di-GMP binding pockets, represents a novel class of c-di-GMP effector proteins that regulate cellular functions in response to c-di-GMP binding.

**Regulatory Significance of DGC Feedback Control**—GGDEF domains are often associated with sensory domains in one- or two-component signaling systems (39, 40). Thus it is reasonable to assume that in most cases DGC activity is controlled by direct signal input through these domains. But why then would a substantial portion of these enzymes also be subject to feedback inhibition? There are several possibilities, which among themselves are not mutually exclusive. Given the anticipated regulatory complexity of the c-di-GMP signaling network (2, 39) and the potentially dramatic changes in cellular physiology and behavior caused by fluctuating levels of c-di-GMP, it is in the cell's best interest to rigorously control the production of the second messenger. Product inhibition of DGCs allows the establishment of precise threshold concentrations of the second messenger, or, in combination with counteracting PDEs, could produce short spikes or even generate oscillations of c-di-GMP. In addition, negative feedback loops have been implicated in neutralizing noise and providing robustness in genetic networks by limiting the range over which the concentrations of the network components fluctuate (41, 42). Similarly, product inhibition of DGCs could contribute to the reduction of stochastic perturbations and increase the stability of the c-di-GMP circuitry by keeping c-di-GMP levels in defined concentration windows. Alternatively, DGC autoregulation may influence the kinetics of c-di-GMP signaling. Mathematical modeling and experimental evidence suggested that negative autoregulation in combination with strong promoters substantially shortens the rise-time of transcription responses (43–45). In analogy, a desired steady-state concentration of c-di-GMP can in principle be achieved by two regulatory designs: (a) a low activity DGC with no product inhibition, and (b) a high activity DGC with built-in negative autoregulation. In cases where circuits have been optimized for fast up-kinetics, design B will be superior. It is plausible that DGCs with or without I-site motifs can be divided into these two kinetically different classes.

This study contributes to the emerging understanding of the c-di-GMP regulatory network in bacteria. The current emphasis lies on the identification of effector molecules, regulatory mechanisms, and processes controlled by c-di-GMP. With the long term goal in mind of approaching a detailed systems-level understanding of c-di-GMP signaling, kinetic parameters of signaling mechanisms will require our particular attention. Our experiments provide an entry point into the kinetic analysis of individual DGCs and the quantitative assessment of the c-di-GMP circuitry.

**Acknowledgments**—We thank Tilman Schirmer for helpful discussions and students of the Microbiology and Immunology Course (Biozentrum of the University of Basel, 2005) and of the Advanced Bacterial Genetics Course (Cold Spring Harbor, 2005) for their help in mutant screening.

## REFERENCES

- Jenal, U. (2004) *Curr. Opin. Microbiol.* **7**, 185–191
- Romling, U., Gomelsky, M., and Galperin, M. Y. (2005) *Mol. Microbiol.* **57**, 629–639
- Brouillette, E., Hyodo, M., Hayakawa, Y., Karaolis, D. K. R., and Malouin, F. (2005) *Antimicrob. Agents Chemother.* **49**, 3109–3113
- Hisert, K. B., MacCoss, M., Shiloh, M. U., Darwin, K. H., Singh, S., Jones, R. A., Ehrst, S., Zhang, Z. Y., Gaffney, B. L., Gandotra, S., Holden, D. W., Murray, D., and Nathan, C. (2005) *Mol. Microbiol.* **56**, 1234–1245
- Kulesekar, H., Lee, V., Brencic, A., Liberati, N., Urbach, J., Miyata, S., Lee, D. G., Neely, A. N., Hyodo, M., Hayakawa, Y., Ausubel, F. M., and Lory, S. (2006) *Proc. Natl. Acad. Sci. U. S. A.* **103**, 2839–2844
- Lestrade, P., Dricot, A., Delrue, R. M., Lambert, C., Martinelli, V., De Bolle, X., Letesson, J. J., and Tibor, A. (2003) *Infect. Immun.* **71**, 7053–7060
- Tischler, A. D., Lee, S. H., and Camilli, A. (2002) *J. Bacteriol.* **184**, 4104–4113
- Chan, C., Paul, R., Samoray, D., Amiot, N. C., Giese, B., Jenal, U., and Schirmer, T. (2004) *Proc. Natl. Acad. Sci. U. S. A.* **101**, 17084–17089
- Ross, P., Weinhouse, H., Aloni, Y., Michaeli, D., Weinbergerohana, P., Mayer, R., Braun, S., Devroom, E., Vandermare, G. A., Vanboom, J. H., and Benziman, M. (1987) *Nature* **325**, 279–281
- Ryjenkov, D. A., Tarutina, M., Moskvina, O. V., and Gomelsky, M. (2005) *J. Bacteriol.* **187**, 1792–1798
- Ross, P., Mayer, R., and Benziman, M. (1991) *Microbiol. Rev.* **55**, 35–58
- Schmidt, A. J., Ryjenkov, D. A., and Gomelsky, M. (2005) *J. Bacteriol.* **187**, 4774–4781
- Tamayo, R., Tischler, A. D., and Camilli, A. (2005) *J. Biol. Chem.* **280**, 33324–33330
- Christen, M., Christen, B., Folcher, M., Schauer, A., and Jenal, U. (2005) *J. Biol. Chem.* **280**, 30829–30837
- Tal, R., Wong, H. C., Calhoun, R., Gelfand, D., Fear, A. L., Volman, G., Mayer, R., Ross, P., Amikam, D., Weinhouse, H., Cohen, A., Sapir, S., Ohana, P., and Benziman, M. (1998) *J. Bacteriol.* **180**, 4416–4425
- Paul, R., Weiser, S., Amiot, N. C., Chan, C., Schirmer, T., Giese, B., and Jenal, U. (2004) *Genes Dev.* **18**, 715–727
- Aldridge, P., Paul, R., Goymer, P., Rainey, P., and Jenal, U. (2003) *Mol. Microbiol.* **47**, 1695–1708
- Kirillina, O., Fetherston, J. D., Bobrov, A. G., Abney, J., and Perry, R. D. (2004) *Mol. Microbiol.* **54**, 75–88
- Aldridge, P., and Jenal, U. (1999) *Mol. Microbiol.* **32**, 379–391
- Hecht, G. B., and Newton, A. (1995) *J. Bacteriol.* **177**, 6223–6229
- Ely, B. (1991) *Methods Enzymol.* **204**, 372–384
- O'Toole, G. A., Pratt, L. A., Watnick, P. I., Newman, D. K., Weaver, V. B., and Kolter, R. (1999) *Methods Enzymol.* **34**, 586–595
- Ochi, Y., Hosoda, S., Hachiya, T., Yoshimura, M., Miyazaki, T., and Kajita, Y. (1981) *Acta Endocrinol.* **98**, 62–67
- Brooks, B. R., Brucoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) *J. Comput. Chem.* **4**, 187–217
- MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiorcikiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998) *J. Phys. Chem. B* **102**, 3586–3616
- Ausmees, N., Mayer, R., Weinhouse, H., Volman, G., Amikam, D., Benziman, M., and Lindberg, M. (2001) *FEMS Microbiol. Lett.* **204**, 163–167
- Simm, R., Fetherston, J. D., Kader, A., Romling, U., and Perry, R. D. (2005) *J. Bacteriol.* **187**, 6816–6823
- García, B., Latasa, C., Solano, C., Portillo, F. G., Gamazo, C., and Lasa, I. (2004) *Mol. Microbiol.* **54**, 264–277
- Zogaj, X., Nitz, M., Rohde, M., Bokranz, W., and Romling, U. (2001) *Mol. Microbiol.* **39**, 1452–1463
- Pei, J., and Grishin, N. V. (2001) *Proteins* **42**, 210–216
- Tucker, C. L., Hurley, J. H., Miller, T. R., and Hurley, J. B. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95**, 5993–5997
- Zhang, G. Y., Liu, Y., Ruoho, A. E., and Hurley, J. H. (1997) *Nature* **388**, 204 (Erratum)

### Diguanylate Cyclase Feedback Control

33. Tesmer, J. J., Sunahara, R. K., Johnson, R. A., Gosselin, G., Gilman, A. G., and Sprang, S. R. (1999) *Science* **285**, 756–760
34. Sunahara, R. K., Beuve, A., Tesmer, J. J. G., Sprang, S. R., Garbers, D. L., and Gilman, A. G. (1998) *J. Biol. Chem.* **273**, 16332–16338
35. Baker, D. A., and Kelly, J. M. (2004) *Mol. Microbiol.* **52**, 1229–1242
36. Hickman, J. W., Tifrea, D. F., and Harwood, C. S. (2005) *Proc. Natl. Acad. Sci. U. S. A.* **102**, 14422–14427
37. Mendez-Ortiz, M. M., Hyodo, M., Hayakawa, Y., and Membrillo-Hernandez, J. (2006) *J. Biol. Chem.* **281**, 8090–8099
38. Simm, R., Morr, M., Kader, A., Nimtz, M., and Romling, U. (2004) *Mol. Microbiol.* **53**, 1123–1134
39. Galperin, M. Y., Nikolskaya, A. N., and Koonin, E. V. (2001) *FEMS Microbiol. Lett.* **204**, 213–214
40. Ulrich, L. E., Koonin, E. V., and Zhulin, I. B. (2005) *Trends Microbiol.* **13**, 52–56
41. Becskei, A., and Serrano, L. (2000) *Nature* **405**, 590–593
42. Gardner, T. S., Cantor, C. R., and Collins, J. J. (2000) *Nature* **403**, 339–342
43. McAdams, H. H., and Arkin, A. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94**, 814–819
44. Rosenfeld, N., Elowitz, M. B., and Alon, U. (2002) *J. Mol. Biol.* **323**, 785–793



### SUPPLEMENTAL MATERIAL:

#### MATERIALS AND METHODS:

*Purification of His-tagged proteins* - *E. coli* BL21 cells carrying the respective expression plasmid were grown in LB medium with ampicillin (100 $\mu$ g/ml) or kanamycin (30 $\mu$ g/ml) and expression was induced by adding IPTG at OD<sub>600</sub> 0.4 to a final concentration of 0.4 mM. After harvesting by centrifugation, cells were resuspended in buffer containing 50 mM Tris-HCl, pH 8.0, 250 mM NaCl, 5 mM  $\beta$ -mercaptoethanol, lysed by passage through a French pressure cell, and the suspension was clarified by centrifugation for 10 min at 5,000 x g. Soluble and insoluble protein fractions were separated by a high-spin centrifugation step (100,000 x g, 1 h). The supernatant was loaded onto Ni-NTA affinity resin (Qiagen), washed with buffer, and eluted with an imidazol-gradient as recommended by the manufacturer. Protein preparations were examined for purity by SDS-PAGE and fractions containing pure protein were pooled and dialyzed for 12 h at 4°C.

#### *Molecular modeling of PleD*

All-atom simulations were carried out using the CHARMM (25) program and the CHARMM22/27 force field (26). The A chain of the X-ray dimer structure (PDB entry: 1W25 (17)) was used. All titratable side chains were generated in their standard protonation state at pH 7. Parameters and partial charges for the non-standard residue c-di-GMP were adopted from the extended CHARMM parameter sets for nucleic acids. The structure of the ligated (intercalated c-di-GMP bound to the I-site) and the unligated protein, to which hydrogen atoms were added, were minimized using a distance-dependent dielectric with  $\epsilon=4$  and a cutoff of 12 Å for non-bonded interactions. 5000 steps of steepest descent minimization were followed by adopted Newton Raphson minimization until a RMS gradient of 10<sup>-7</sup> kcal/mol·Å was reached. Such a threshold is found to be sufficient for normal mode calculations (49). Normal modes were calculated with the diagonalization in a mixed basis (DIMB) method, as implemented in CHARMM. The DIMB method is an approximate scheme retaining the full atomistic description of the protein, where the Hessian is approximated iteratively. The total number of basis functions was 153 and cumulated displacements were calculated for  $T = 300$  K.

For ligated PleD motion is suppressed at L( $\beta$ 1, $\alpha$ 1) (res.10-12), L( $\beta$ 3, $\alpha$ 3) domain REC1, the C-terminal end of  $\alpha$ 3 (res. 220-224) of domain REC2, the unstructured linker between REC2 and GGDEF domain (res. 282-284), the residues forming the A-site (res. 352), L( $\alpha$ 2, $\beta$ 2) (res. 357-

360, I-site), L( $\beta$ 2, $\beta$ 3) (res. 367-373, A-site) and at the C-terminal end of  $\alpha$ 3 (res. 396-398) of the GGDEF domain. By contrast upon ligand binding mobility increases for  $\alpha$ 1 (res. 24),  $\alpha$ 4 (res. 96-99) of domain REC1, residues (res. 149, 175), L( $\beta$ 2, $\alpha$ 2) (res. 205-207), L( $\beta$ 5, $\alpha$ 5) (res. 254-257) of domain REC2 and residues L( $\beta$ 3', $\beta$ 3'') (res. 404-407) and L( $\beta$ 4, $\alpha$ 4) (res. 422-424) of the GGDEF domain.

*Primer list*

The following primers were used: #1006, ACA CGC TAC ATA TGA AAA TCT CAG GCG CCC GGA C; #1007, ACT CTC GAG AGC GCT CCT GCG CTT; #1129, CAA GCG GCT GCA GGC CAA TGT GAT CGT CGG CCG CAT GGG TGG TGA; #670, TGC TAG TTA TTG CTC AGC GG; #1006 ACA CGC TAC ATA TGA AAA TCT CAG GCG CCC GGA C; #1130, CAA GCG GCT GCA GGC CAA TGT GCG CGA AAG CGA CAT CGT CGG CCG CAT GGG TGG TGA; #1132, CAC ATT GGC CTG CAG CCG CTT GGC GAC; #1131, CAA GCG GCT GCA GGC CAA TGT GNN NNN NNN NAT CGT CGG CCG CAT GGG TGG TGA.

FIGURE LEGENDS:

**Figure S1: Separation of peptides yielded from tryptic digest of PleD in the presence (red chromatogram) or absence of c-di-GMP (black chromatogram) on a C18 column.** Peaks identified by ESI-MS: c-di-GMP  $m/z$  691,  $t_R$  7.70 min, T47 (amino acids 354-359)  $m/z$  659.3  $t_R$  25.64 min. T49 (amino acids 367-386)  $m/z$  2167.7  $t_R$  47.73 min.

**Figure S2: Normal modes of PleD I-site and A-site residues.** The displacements for each mode of the ligated and unligated structures are shown in Å for the residues of the REC2 domain (green) and the GGDEF domain (red). Insertion of intercalated c-di-GMP in the I-site quenches motion in both the I-site (R359-D362, R390) and the A-site (G368-E371), suggesting that the two sites are dynamically coupled.

**Figure S3: Representation of the PleD protein (blue: REC1, green: REC2, red: DGC) with c-di-GMP bound to the I-site.**  $C\alpha$ -atoms at positions of considerable changes in flexibility upon ligand binding are shown as spheres; reduced flexibility (yellow) and enhanced flexibility (black). Note that binding of c-di-GMP at the I-site (I) affects mobility not only in the I-site, but also in other regions of the protein, e.g. A-site (A), phosphorylation site (P) and dimer interface.

**Figure S4: Alignment of I- and A-site sequence of biochemically characterized diguanylate cyclases.** I-site residues (RXXD) are underlined in green and A-site residues (GGDEF) are underlined in yellow.



## FIGURE LEGENDS:

**Figure S1: Separation of peptides yielded from tryptic digest of PleD in the presence (red chromatogram) or absence of c-di-GMP (black chromatogram) on a C18 column.** Peaks identified by ESI-MS: c-di-GMP  $m/z$  691,  $t_R$  7.70 min, T47 (amino acids 354-359)  $m/z$  659.3  $t_R$  25.64 min. T49 (amino acids 367-386)  $m/z$  2167.7  $t_R$  47.73 min.

**Figure S2: Normal modes of PleD I-site and A-site residues.** The displacements for each mode of the ligated and unligated structures are shown in Å for the residues of the REC2 domain (green) and the GGDEF domain (red). Insertion of intercalated c-di-GMP in the I-site quenches motion in both the I-site (R359-D362, R390) and the A-site (G368-E371), suggesting that the two sites are dynamically coupled.

**Figure S3: Representation of the PleD protein (blue: REC1, green: REC2, red: DGC) with c-di-GMP bound to the I-site.**  $C\alpha$ -atoms at positions of considerable changes in flexibility upon ligand binding are shown as spheres; reduced flexibility (yellow) and enhanced flexibility (black). Note that binding of c-di-GMP at the I-site (I) affects mobility not only in the I-site, but also in other regions of the protein, e.g. A-site (A), phosphorylation site (P) and dimer interface.

**Figure S4: Alignment of I- and A-site sequence of biochemically characterized diguanylate cyclases.** I-site residues (RXXD) are underlined in green and A-site residues (GGDEF) are underlined in yellow.

Figure S1

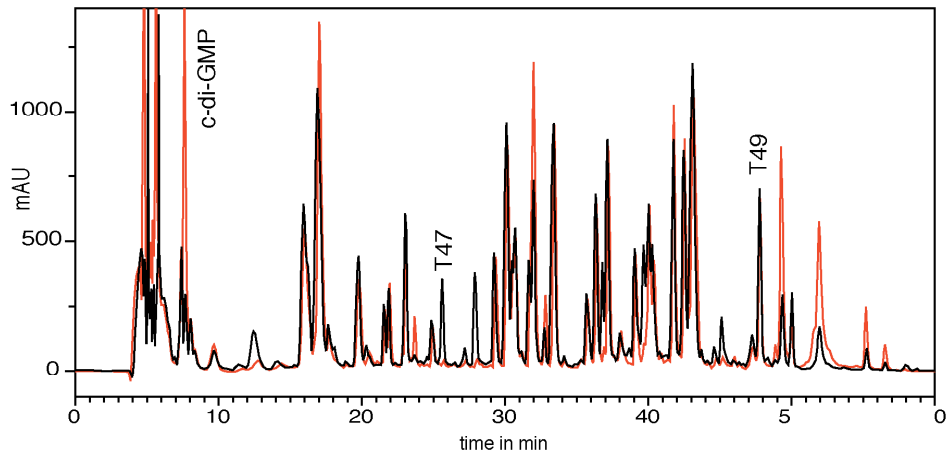


Figure S2

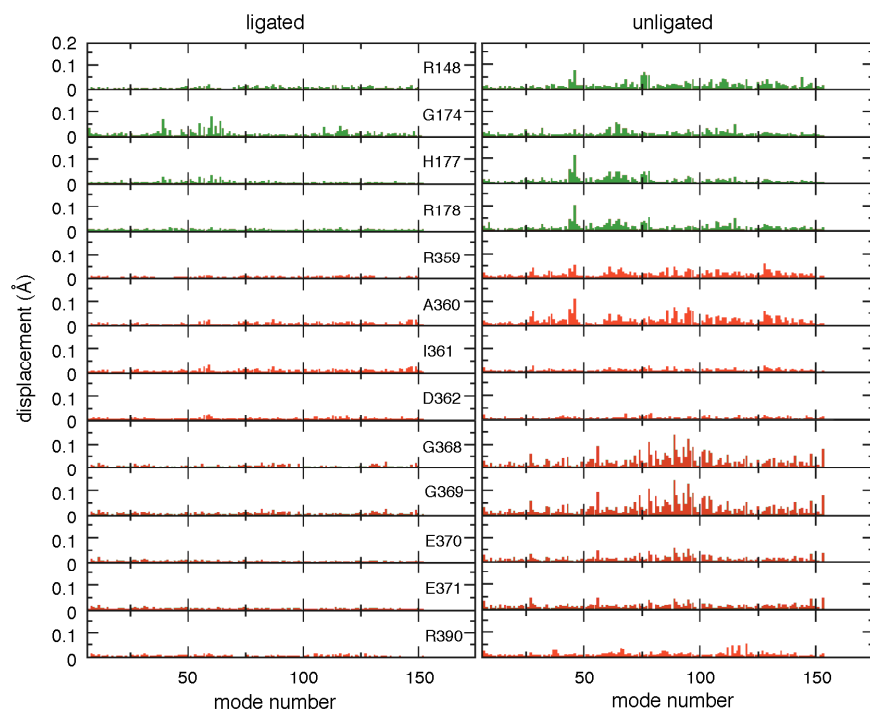


Figure S3

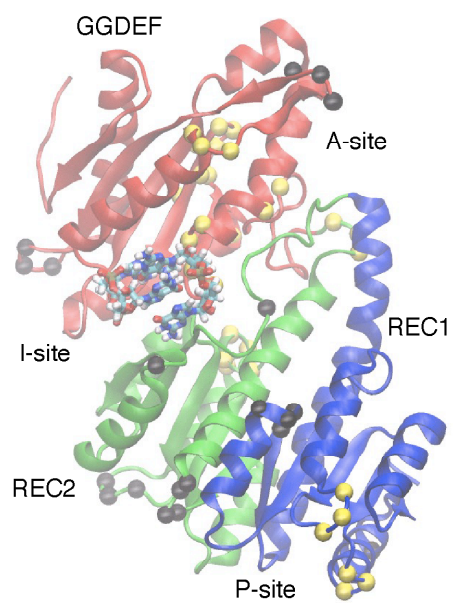


Figure S4

Gene	Sequence	Reference
BORBU_BB0419	LKYKIDVARYGGEEFI	(Ryjenkov et al, 2005)
CAUCR_DgcA	VRESDIVGRMGDEFA	- this study
CAUCR_PleD	VRAIDLPCRYGGEEFV	(Paul et al, 2004)
DEIRA_DRB0044	LPGGASLYRVGGDEEV	(Ryjenkov et al, 2005)
ECOLI_YddV	VRSSDYVFRYGGDEFI	(Mendez-Ortiz, 2006)
ECOLI_YeaP	QQNGEVIGRLGGDEFL	(Ryjenkov et al, 2005)
PSEAE_PA1107	TRSSDSVARLGGEEFL	(Kulesekara et al, 2006)
PSEAE_PA1120	LRESDLVARLGGDEFA	(Kulesekara et al, 2006)
PSEAE_PA1727	VRAQDTIARLGGDEEV	(Kulesekara et al, 2006)
PSEAE_PA2870	LREVDLLGRLGGEEFA	(Kulesekara et al, 2006)
PSEAE_PA3343	RRFLDMAVRLGGEEFA	(Kulesekara et al, 2006)
PSEAE_WspR	SRSSDLAARYGGEEFA	(Hickman et al, 2005)
PSEPA_PA0847	LKQFDKAYRLGGDEFA	(Kulesekara et al, 2006)
RHOSP_Rsp3513	IGPADALGRIGGEEFA	(Ryjenkov et al, 2005)
SALTY_AdrA	LRGSDIIGRFGGDEFA	(Simm et al, 2005)
SYNY3_Slr1143	LRSYDILGRWGGDEEM	(Ryjenkov et al, 2005)
THEMA_TM1163	VRESDLVFRYGGDEFL	(Ryjenkov et al, 2005)
VIBCH_VCA096	CRDGVTAARYGGEEFA	(Ryjenkov et al, 2005)
YERPS_HmsT	VRSRDIVVRYGGEEFL	(Simm et al, 2005)
Consensus	r d R GG#EF	

## 5.2 All-atom simulations of structures and energetics of c-di-GMP bound and free PleD

F. F.-F. Schmid and M. Meuwly  
*J. Mol. Biol.* (2007) 374:1270-1285

**JMB**Available online at [www.sciencedirect.com](http://www.sciencedirect.com)
 ScienceDirect


## All-atom Simulations of Structures and Energetics of c-di-GMP-bound and free PleD

Franziska F.-F. Schmid and Markus Meuwly\*

Department of Chemistry,  
University of Basel,  
Klingelbergstrasse 80,  
CH-4056 Basel, Switzerland

Received 14 May 2007;  
received in revised form  
1 September 2007;  
accepted 22 September 2007  
Available online  
29 September 2007

Cyclic diguanosine monophosphate is a bacterial second messenger involved in a lifestyle switch from single cells to biofilm formation. Atomistic simulations are used to characterize inhibited diguanylate cyclase (DGC) PleD with emphasis on the feedback inhibition mechanism. Normal-mode calculations show a rigidification particularly in both the inhibition site and the active site of the protein upon ligand binding. Extensive molecular dynamics simulations in explicit solvent and analysis of the dynamical cross-correlation maps suggest two distinct coupling pathways between the active and the inhibition site: direct information transfer either through the  $\beta$ -strands  $\beta 2$  and  $\beta 3$  of the DGC domain (pathway I) or *via* the disordered regions connecting domains D2 and DGC (pathway II). In addition, dynamical cross-correlation maps show differences in the correlation between neighboring domains upon ligand binding and upon the point mutation R390A. The correlated motions between domains D1 and D2, which form the dimerization interface, are stronger for free PleD. Complementary to the experimentally observed short-range interactions in ligated PleD, the present work also characterizes the long-range, delocalized interactions between domains that are important for understanding activation and allosteric control of the protein. Based on the results, experimental characterization of the point mutant R353 and of the double mutant N357/H394 is proposed to differentiate between pathways I and II.

© 2007 Elsevier Ltd. All rights reserved.

Edited by M. Levitt

**Keywords:** MD simulations; second messenger c-di-GMP; diguanylate cyclase; product inhibition; protein function

### Introduction

Cyclic diguanosine monophosphate (c-di-GMP) has recently been identified as a second messenger in bacteria important for multicellular behavior and persistence. Bacterial biofilm formation is involved in life-threatening infectious diseases, such as cystic fibrosis, or the colonization of medical devices. Cellular levels of c-di-GMP are controlled through diguanylate cyclases (DGCs) and c-di-GMP-specific

phosphodiesterases.<sup>1</sup> DGCs convert two molecules of guanosine triphosphate (GTP) into c-di-GMP, while phosphodiesterases cleave the cyclic compound into the linear dinucleotide pGpG.<sup>2</sup> Recently, an involvement of c-di-GMP in regulation of host-pathogen interactions including antioxidant defense and cytotoxicity have been proposed.<sup>3</sup> Thus, inhibition of DGCs is a potential target to combat biofilm-related infections.

Genetic studies support the notion that DGCs control bacterial growth on surfaces by regulating motility and cellular adhesion.<sup>4</sup> This rests to some extent on investigations of PleD from *Caulobacter crescentus*, a protein that efficiently catalyzes the conversion of GTP into c-di-GMP.<sup>5</sup> PleD is a multi-domain regulator protein with a C-terminal DGC domain that serves as an output module.<sup>5,6</sup> The DGC domain (also called GGDEF or DUF1) is highly conserved in many bacterial species and corresponds to one of the largest protein families in the database. The DGC domain of PleD is associated

\*Corresponding author. E-mail address: [m.meuwly@unibas.ch](mailto:m.meuwly@unibas.ch).

Abbreviations used: ABNR, adopted Newton–Raphson; BNM, block normal mode; c-di-GMP, cyclic diguanosine monophosphate; DCCM, dynamical cross-correlation map; DIMB, diagonalization in a mixed basis; DGC, diguanylate cyclase; GTP, guanosine triphosphate; MD, molecular dynamics; NMs, normal modes; NMA, normal-mode analysis; RMSE, root-mean-square fluctuation.

with two N-terminal CheY-like receiver domains (D1 and D2). The PleD DGC is activated by specific phosphorylation of the first receiver domain (D1).<sup>7</sup>

Very recently, the structure of PleD complexed with c-di-GMP has been successfully determined using X-ray crystallography.<sup>8</sup> This structure identified a potential allosteric inhibition site (I-site) with two mutually intercalated c-di-GMP molecules tightly bound to the DGC/D2 domain interface.<sup>8</sup>

Structure/function analysis of PleD and DgcA, a related DGC, has confirmed the role of the I-site in feedback inhibition control.<sup>2</sup> Biochemical, genetic and structural analysis of CheY had indicated that phosphorylation induces a structural change upon activation.<sup>9–13</sup> However, it is unclear how much the phosphorylation-induced conformational changes contribute to the observed stabilization of the active state over the inactive state in CheY.<sup>9</sup> It is interesting to note that the structural difference between the active and the inactive form of CheY is quite small (on average below 1 Å) and restricted to a small region of the protein (around residues 90 and 113, which are loop regions).<sup>10</sup>

So far, little is known about how DGCs work at a molecular level. For example, the mechanism of DGC feedback inhibition, mentioned above, is not known. For this, structural information on the protein and its ligand is required. Possible modes of action have been suggested on the basis of the X-ray structure and general features based on structure–function relationships.<sup>8</sup> One possibility is that the binding of c-di-GMP results in domain intercalation and immobilization of the DGC and D2 domains. Alternatively, upon I-site occupation, the dimerization interface may be altered, preventing dimer formation and thus inhibiting DGC activity.<sup>8</sup> The latter proposal is consistent with the observation that PleD has to form a dimer for activation. Finally, it is possible that binding of c-di-GMP to the I-site leads to intradomain rearrangements that would alter the catalytic properties of the active site.<sup>2</sup>

The present study investigates inhibited monomeric PleD with intercalated c-di-GMP present in the inhibition site in view of a possible biological mechanism for allosteric control. Traditionally, protein function has been inferred from its structure, and possibly additional information (such as kinetic and spectroscopic data). However, contrary to the static view of protein function, the view emerged that protein dynamics contributes considerably to function.<sup>14</sup> One example are collective, usually low frequency vibrations (so-called functional modes) that can play important roles in conformational transitions relevant to protein function.<sup>14</sup>

## Results

In the following, results are presented for structural properties (energy-minimized structures and *B*-factors), which are compared to data from experiments, for the exploration around the energy-minimized structures (normal modes), and for the

dynamics [through molecular dynamics (MD) simulations and analysis of the correlated motions]. Results are first presented for the ligated native protein, for which experimental data are available. Then, changes compared to ligated PleD are discussed upon removing the ligand. To describe structural features the following nomenclature is used:  $\alpha X$  for  $\alpha$ -helices,  $\beta X$  for  $\beta$ -strands where  $X$  is a number. Loop regions are written as  $L(x,y)$  where  $x$  and  $y$  are the flanking secondary-structure elements.

### Structures and validation of the ligated and unligated protein

#### Ligated PleD

Figure 1 gives an overview of the structure of one monomer of ligated PleD. The DGC domain (top) contains the active site (cyan), whereas the inhibition site (orange) is located between domains DGC and D2. The active site (A-site production site for c-di-GMP) is located in the region of the  $\alpha 1$  and  $\alpha 2$  helices with the conserved GGDEF motif in the loop  $L(\beta 2, \beta 3)$  of the DGC domain (residues G368, G369, E370, E371).<sup>5</sup>

Optimized structures were calculated as described in Materials and Methods for ligated PleD starting from the X-ray structure. The final RMS difference between the calculated and experimentally determined structure is 0.97 Å (constrained) and 1.77 Å (free) for the backbone atoms (residues 2–454), respectively. Such an RMSD is quite typical for force field optimizations of medium to large proteins using a similar minimization protocol.<sup>15,16</sup> Normal-mode calculations [block normal-mode (BNM) method] using the X-ray structure of inhibited PleD yielded 149 negative normal-mode frequencies and hence this is not a minimum-energy structure. Regions showing considerable rearrangements during structural optimization also exhibit large fluctuations in the imaginary modes, namely, the unstructured parts connecting the domains D1–D2 and D2–DGC.

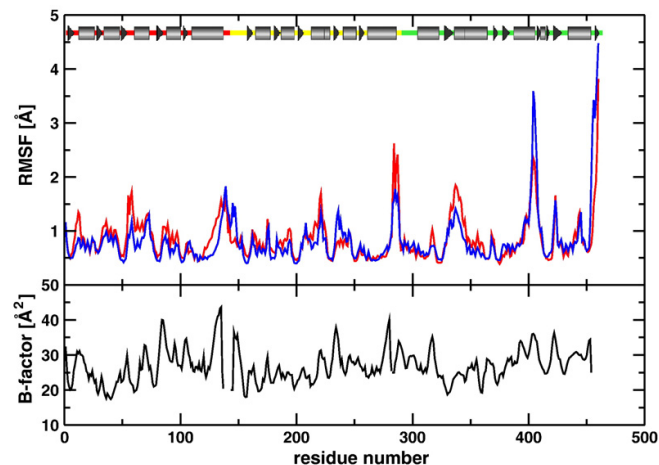
#### *B*-factors for ligated PleD

The MD simulations of ligated PleD (for details, see below) also allow one to calculate root-mean-square fluctuations (RMSFs), which are a measure of the flexibility and fluctuations around the average protein structure and which are related to measured *B*-factors. They are calculated over the last 1.25 ns of the MD simulations. Figure 2 shows that, in general, the location of the flexible regions—characterized by larger *B*-factors and RMSF—and the more rigid parts of the protein agree quite well. In regions where no electron density was found in the experiment (residues 138–146 and residues 283–285), the calculated RMSFs are large. These parts of the protein correspond to linking regions between the different domains.

Experimental *B*-factors and RMSFs differ in the region of residues 325 to 345, which are adjacent to







(residues 54–57, 104–108); disordered region and part of  $\alpha 3$  of D2 (residues 144–147, 219–221); L( $\beta 4$ ,  $\alpha 4$ ) of DGC (residues 420–424). In general, the pattern from the experiment is well reproduced by the simulations (see, e.g., the region around residues 400 and 420).

which c-di-GMP was removed. The final RMS difference is 0.98 Å (constrained) and 1.88 Å (free) for the backbone atoms (residues 2–454), similar to the RMSD found for ligated PleD. Ligand removal leads to structural differences mainly in the DGC domain and the unstructured regions connecting the different domains. BNM calculations yield two (constrained) and three (free) slightly negative normal modes ( $\leq 0.001 \text{ cm}^{-1}$ ), respectively, which shows that the structures correspond to real minima on the potential energy surface. However, it is quite likely that the structures investigated do not correspond to a low-energy structure (see also Discussion and Conclusions).

#### Fluctuations and change of flexibility upon complexation

To characterize the influence of ligand binding on the flexibility of the protein, the C $^{\alpha}$  RMSFs of ligated and unligated PleD are compared (see Fig. 2). Differences are found in the following regions: RMSF values for unligated PleD are higher for L( $\beta 1$ ,  $\alpha 1$ ) (residues 8–15), L( $\beta 3$ ,  $\alpha 3$ ) and neighboring  $\alpha 3$  (residues 54–70) and the C-terminal half of  $\alpha 5$  (residues 120–138) of domain D1,  $\alpha 3$  (residues 210–225) of domain D2, the linking region between D2 and DGC (residues 283–289) and L( $\alpha 0$ ,  $\beta 1$ ) (residues 315–319) of the domain DGC. In contrast, RMSF values for unligated PleD are lower for the linking region between D1 and D2 (residues 138–150), L( $\beta 4$ ,  $\alpha 4$ ) (residues 234–239) of domain D2 and L( $\beta 3'$ ,  $\beta 3''$ ) (residues 402–408) of domain DGC.

To summarize, it is found that the structure and B-factors of ligated PleD agree well with experimental findings, while for unligated PleD the structure can be further relaxed upon removing the ligand. This is consistent with experiments, which find that the structures of ligated and unligated PleD differ. Comparing the RMSFs between ligated and

**Fig. 2.** B-factors versus calculated RMSF. Experimental B-factors and RMSF (in angstroms) from the MD simulations for ligated (black and blue) and unligated PleD (red). The secondary-structure elements are indicated on top: arrows for  $\beta$ -strands and rectangular boxes for helices. Domains D1, D2 and DGC are color coded in red, yellow and green, respectively. It should be noted that in regions where no electron density was found experimentally (residues 138–146, 283–285) the B-factor in the PDB file was arbitrarily set to 20. Prominent differences between the calculated B-factors for ligated and unligated PleD occur mostly in loop regions: L( $\beta 3$ ,  $\alpha 3$ ) and L( $\beta 5$ ,  $\alpha 5$ ) of D1

unligated PleD indicate subtle but potentially important differences in the local flexibilities of secondary-structure elements also away from the I site and in the linker regions between domains. This point will be further discussed in the MD simulations.

#### Normal-mode analysis

Normal modes (NMs), that is, excursions of a molecule around a minimum-energy structure, are usually calculated from diagonalizing the  $3N \times 3N$  Hessian matrix (matrix of second derivatives), where  $N$  is the number of atoms of the molecule.<sup>18–20</sup> For free ( $N=7140$ ) and complexed ( $N=7276$ ) PleD the number is too large for direct diagonalization of the entire Hessian. Consequently, approximate methods such as diagonalization in a mixed basis (DIMB) or BNM are used for calculating vibrational frequencies. DIMB retains the full atomistic description of the protein, whereas BNM is a coarse-grained model (see Materials and Methods). First, the results of DIMB are described and then compared to BNM calculations.

To validate DIMB for the present work, the full set of NMs was calculated for isolated c-di-GMP and compared with DIMB using  $m=103$  basis functions. This leads to 97 nonzero NMs from DIMB. For these 97 NMs, no difference was found between DIMB and frequencies based on the diagonalization of the full Hessian matrix. These results give confidence that DIMB applied to the free and ligated structure of PleD is meaningful, especially for the low-frequency modes. Convergence studies were also carried out for ligated and unligated PleD with  $m$  ranging from 50 to 300 (53, 103, 153, 175, 200, 225, 250, 275 and 300). Increasing  $m$  also increases the number of atoms in a particular block and the normal-mode frequencies shift to lower values until they converge for  $m > 275$ . To further assess the influence of the basis set size used with DIMB,

RMSFs of the normal modes at  $T=300$  K were calculated for  $m=53, 103$  and  $153$ . The RMS fluctuations from normal mode analysis (NMA) agree qualitatively with those obtained from the MD simulations and are shown in Fig. 3. Furthermore, identification of the flexible regions is independent of the basis set size. It should be mentioned, however, that NMA explores the region around the minimum of the potential energy surface, while MD simulations allow structural rearrangements. In addition, the MD simulations are carried out in explicit solvent, while for the NMA a distance-dependent dielectric is used.

As an alternative, a coarse-grained method (BNM) was used to calculate the normal modes. This analysis provides a comparison of NMs based on DIMB with a method that allows one to calculate the majority of low-frequency NMs. Because BNM contains one phenomenological coupling constant between neighboring residues and not explicit masses,<sup>21</sup> the frequencies are only determined up to an approximate multiplicative factor and absolute frequencies cannot be directly compared to a DIMB calculation. However, it is possible to characterize the large-scale motions of the protein.

#### Overview of the normal modes

Figure 4a shows the frequencies as a function of the mode number calculated with DIMB (basis set size 153) (solid line) and BNM (dashed line). It is found that the BNM frequencies are lower than the ones from DIMB, whereas the density of states is comparable. For both methods, frequencies for most normal modes are slightly higher for ligated PleD (blue) than for the unligated protein (red). This is a hint that ligand binding influences the flexibility of the protein such that the complex is more rigid. The mode density distribution for both DIMB and BNM is shown in Fig. 4b and c. When comparing the percentage of modes with frequencies

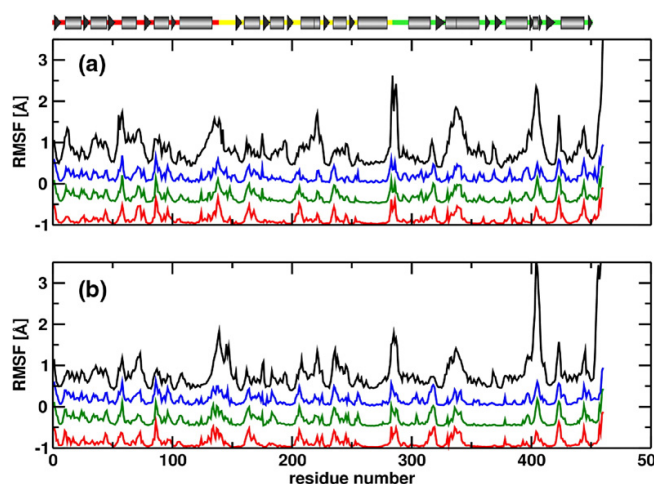
below  $10 \text{ cm}^{-1}$ , one finds 7.5% and 7.9% for unligated PleD, which is higher than 5.4% and 7.7% for the complex system, respectively. This distribution difference is more pronounced in the case of DIMB. The effect of ligand binding on the protein flexibility is analyzed in more detail for residues of the I- and A-site.

#### Amplitudes per mode for I- and A-site residues

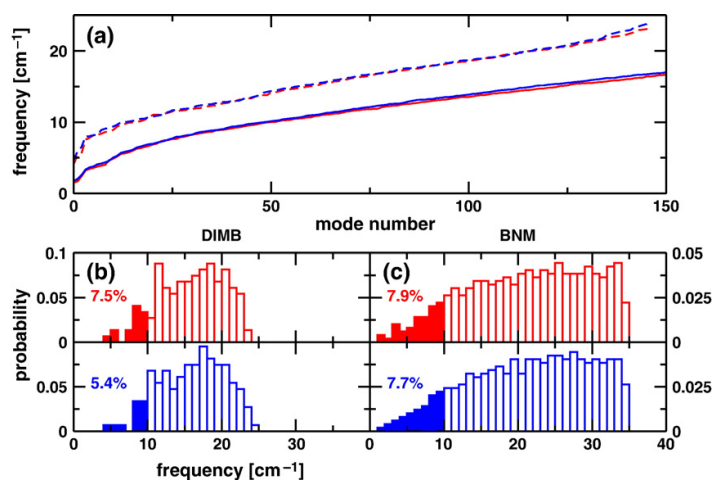
From an NMA it is also possible to investigate whether the magnitude of motion of a particular residue is damped or enhanced upon ligand binding and which residues and parts of the protein participate in a particular NM. A considerable displacement for a given mode suggests that this residue is strongly involved in the mode. The displacements for the residues in the I- and A-site for each of the 147 nonzero modes from DIMB are shown in Fig. 5. Residues 368 to 371 are involved in binding of GTP in the A site, whereas residues 148, 174, 177–178, 359–362 and 390 were identified to be important for feedback inhibition with c-di-GMP.<sup>8</sup>

For ligated and unligated PleD,  $C^\alpha$  displacements for each NM were calculated at  $T=300$  K. It is found that the displacements for the A-site residues are considerably larger for free PleD than for the ligated protein. In particular, motion for modes 75 to 100 is greatly enhanced for unligated PleD. The effect is less pronounced for the I-site residues, albeit R359 and A360 are still more mobile for the free system. NMA suggests that ligand insertion reduces the flexibility of the I- and A-sites.

To ensure that the reduced flexibility in both the I- and A-sites is indeed due to ligand binding and not an artifact of the NMA, displacements for distant residues were calculated. These control residues (not shown) display random increase/decrease of displacements upon ligand binding. When the basis set size is increased to 300 (with concomitant increase of the number of atoms per block), the difference in



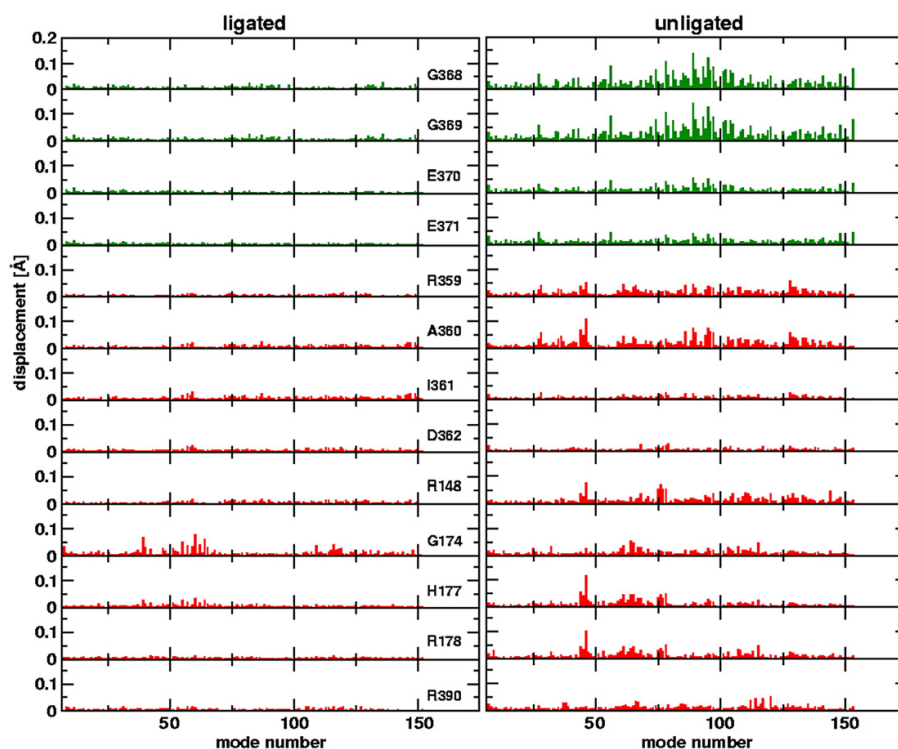
**Fig. 3.** Residue mobility from MD simulations and normal mode calculations. Comparison between RMSFs from the MD simulations and from normal mode calculations (DIMB) with different basis set sizes for (a) unligated and (b) ligated PleD. The following color code is used: MD, black; DIMB with basis set size 153, blue; 103, green; and 53, red. The secondary-structure elements are indicated on top: arrows for  $\beta$ -strands and rectangular boxes for helices. Domains D1, D2 and DGC are color coded in red, yellow and green, respectively. The shape of the RMSF obtained from the MD simulation and the NMA are very similar. There is an improvement if a higher number of NMs is included (see, e.g., around residue 175).



**Fig. 4.** Normal mode frequencies. (a) The normal mode frequencies obtained from DIMB with  $m=153$  basis functions (dashed line) and BNM calculation (continuous line). The absolute wave numbers from the BNM calculations are lower but the shape of the curves is comparable. (b) and (c) Density distribution of the frequencies from DIMB and BNM, respectively. In all plots, ligand binding shifts most normal modes to slightly higher frequencies. The number of modes with frequencies below  $10 \text{ cm}^{-1}$  is higher in the unligated PleD (7.5% versus 5.4% for DIMB and 7.9% versus 7.7% for BNM) than in the complex structure. The difference in mobility is more pronounced for DIMB. In all plots, ligated PleD is blue and unligated PleD is red.

mobility between ligated and unligated PleD is less pronounced. However, ligand binding still reduces the mobility of R359, A360, I361, E362, R148, H177, R178 and R390.

The NMs also allow one to study the influence of ligand binding on the mobility of residues that form the I- and A-sites (see Fig. 5). Comparing the residual displacements in ligated and unligated



**Fig. 5.** Influence of ligand binding on active and inhibition site residues.  $C^\alpha$  displacements calculated from normal-mode calculations (DIMB) at  $T=300 \text{ K}$  for I- (red) and A-site (green) residues, respectively. Ligand binding in the inhibition site decreases the mobility of the residues involved in binding. However, the mobility of residues in the distant A-site is also considerably damped.

PleD, we find that the most striking feature is the suppression of motion upon ligand binding in both the I- and A-sites. The influence of ligand binding on the dynamics of residues in both sites can be attributed to dynamical coupling. One interesting point is to understand how the two sites are coupled in detail. Looking at the structure, we can envisage two possibilities for information transfer: localized coupling involving the short  $\beta$ -strand  $\beta 2$  that connects the I- and the A-sites. Alternatively, the coupling may be distributed over the entire protein. To get information about the coupling mechanism, for example, through which structure elements coupling occurs, the NMs that involve residues from the I- and A-sites are characterized. In the following, modes 27, 56 and 78 and the range from 89 to 98 calculated with DIMB (basis set size 153) will be considered more closely. Modes 27, 78 and those from 89 to 98 involve both I- and A-site residues, whereas in mode 56 only A-site residues participate.

#### Modes 89 to 98

$C^\alpha$  atoms of A- (residues G368 to E371) and I-site (R359, A360) residues have displacements of up to 0.1 Å for these modes and the modes are delocalized over the entire protein. The largest amplitudes are found in the DGC domain. Helices  $\alpha 3$  and  $\alpha 1$  are anticorrelated to  $L(\beta 2, \beta 3)$  (A-site) and  $\beta 3''$ . The loops  $L(\beta 2, \beta 3)$  and  $L(\alpha 2, \beta 2)$  (A- and I-site loops) are connected *via* the short  $\beta$ -strand  $\beta 2$ , which mediates motion. Structural elements forming the I-site pocket are correlated with the I-site loop. Overall, a screw-like motion of the helices  $\alpha 1$ ,  $\alpha 2$  and  $\alpha 3$  in the DGC domain translates  $\beta 2$  along the strand direction.

#### Modes 27 and 78

Considerable displacements in both sites are found for modes 27 and 78. In mode 27 the I- and A-site loops, connected *via* the  $\beta$ -strand  $\beta 2$ , move together in the same direction. The helix  $\alpha 0$  seems to displace the  $\beta$ -strand  $\beta 2$ . However, this specific mode leads to major movement of domain D1 without the extended helix  $\alpha 5$  as a unit and in the DGC domain  $\alpha 0$ ,  $L(\beta 4, \alpha 4)$ ,  $\beta 3'$  to  $\beta 3''$  and  $L(\alpha 1, \alpha 2)$  with parts of the neighboring helices. The D2 domain remains stationary. In mode 78, substantial motion is found in the I-site region, especially the disordered linking regions and the C-terminal part of  $\alpha 5$  of D1. The loops  $L(\alpha 2, \beta 2)$  and  $L(\beta 2, \beta 3)$  (I- and A-site loop) move in the opposite direction, as found for modes 89 to 98. Other parts of the protein that display motion in this mode are  $\alpha 4$  to  $\beta 5$  and both ends of  $\alpha 0$  of DGC,  $\beta 2$  to  $\alpha 2$  of D2 and  $L(\beta 3, \alpha 3)$  of D1.

#### Mode 56

Contrary to the above, mode 56 involves only motion of A-site residues. Further movements are found in the DGC domain,  $\alpha 1$  of D2 and  $\alpha 2$  and the

C-terminal half of  $\alpha 5$  of D1.  $L(\beta 2, \beta 3)$  moves together with  $\alpha 1$  and  $\alpha 4$  of DGC. The motion of the N-terminal part of  $\alpha 3$  is in the opposite direction. The C-terminal half of  $\alpha 5$  of D1 and  $\beta 0$  of DGC also move in the direction opposite that of the A-site loop.

In summary, the NMA suggests that the I- and A-sites in DGC are dynamically coupled *via* the short  $\beta$ -strand  $\beta 2$  that itself is influenced from motion of the nearby helices. In addition, dynamical coupling between the two sites is not restricted to a single functional mode but is present in several low-frequency modes.

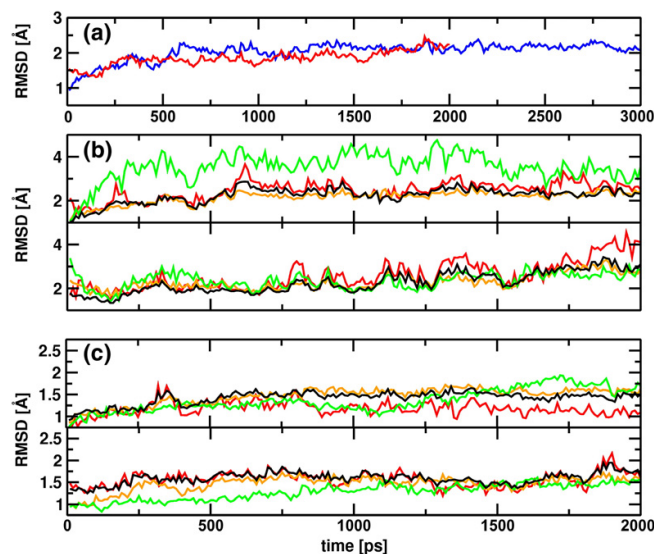
#### Molecular dynamics simulations of native, complexed PleD

Complementary to NMA, MD simulations provide additional information about collective motions within a protein, including a realistic solvation environment. One further important difference is that NMA is sensitive to motion around the optimized structure, whereas MD simulations also capture anharmonic contributions and allow, for example, crossing of torsional barriers. Representative of all trajectories, the MD simulations for ligated PleD are described in more detail, because for this system a well-defined starting structure is available from experiment.

During the MD simulations all secondary-structure elements are preserved. The average RMSD of all  $C^\alpha$  atoms, compared to the X-ray structure, is 2.0 Å (see Fig. 6a). Surprisingly, the RMSD for the complex is higher than for the unligated protein. This is mainly due to a reorientation of the DGC domain with respect to domains D1 and D2 (see Figs. 6b and 7c). At longer simulation times this feature becomes less pronounced and the RMSD for both systems is  $\approx 2.0$  Å. In particular, DGC domain reorientation is reduced for trajectory III (see also Fig. 3 in Supplementary Data). To distinguish between structural changes within the domains and rearrangement between the three domains, snapshots taken every 10 ps were reoriented with respect to each domain (D1, D2 and DGC). Then, average RMSDs were calculated for all  $C^\alpha$  atoms of PleD (see Fig. 6b) and for the  $C^\alpha$  atoms of each domain, respectively (see Fig. 6c). Comparing the average RMSDs for  $C^\alpha$  atoms within each domain after superposition on the domain considered gives information about intradomain structural changes.

RMSDs for the  $C^\alpha$  atoms in domains D1, D2 and DGC are 1.2, 1.5 and 1.4 Å, respectively. Interdomain changes can be assessed from the difference in RMSD of all  $C^\alpha$  atoms after reorientation on a single domain (see Fig. 6b). The RMSD considering domains D1 and D2 give similar values (2.5 Å for D1 and 2.2 Å for D2). However, the catalytically active domain DGC is oriented differently with respect to the other domains, which is also reflected in an RMSD of 3.5 Å (see Fig. 7c).

To represent global motions of the protein, snapshots taken every 250 ps are superimposed (see



**Fig. 6.** RMSD during the MD simulation. RMSDs for  $C^\alpha$  atoms with respect to the X-ray structure of ligated PleD from the MD simulations in explicit solvent. (a) The RMSD calculated for all  $C^\alpha$  atoms reaches a constant value after  $\approx 250$  ps for unligated PleD (red) and  $\approx 500$  ps for ligated PleD (blue). To ensure the full equilibration of the system, the simulation of ligated PleD was extended to 3.0 ns. (b) RMSD for all  $C^\alpha$  atoms after superposition on domains D1 (red), D2 (orange), DGC (green) and D1/D2 (black), respectively, are shown in the middle panel (top, ligated; bottom, unligated). For the complex, superposition on the domain DGC leads to a significantly larger RMSD compared to the other cases. (c) RMSD calculated for each domain separately (same color code; top, ligated; bottom, unligated).

Fig. 7) onto the backbone atoms. The most flexible parts of the protein are the region  $\beta 3'/\beta 3''$  and the C-terminus of the DGC domain. In addition, for the complex system, one observes the reorientation of the DGC domain with respect to domains D1 and D2 (see Fig. 7c). This is also reflected in the RMSD of all  $C^\alpha$  atoms after reorientation on domain DGC (3.5 Å versus 2.4 Å for ligated and unligated PleD, respectively). The D1/D2 interface is also rearranged.

For unligated PleD, major shifts of secondary-structure elements are observed in the DGC and D1 domain (see Fig. 7a) compared to the X-ray structure. The helix  $\alpha 0$  of the DGC domain moves towards the unoccupied I-site pocket, and domain D1, except for the extended helix  $\alpha 5$ , is moved along the axis of the  $\alpha$ -helices  $\alpha 2$  and  $\alpha 3$ . The major rearrangement—the RMSD for  $C^\alpha$  atoms reaches about 2.4 Å—is the reorientation of domain DGC relative to domains D1 and D2 for the complex. In all cases, the X-ray structure was used for comparison. Therefore, one cannot exclude artifacts from crystal packing and it has to be kept in mind that simulations were carried out for the PleD monomer, whereas the asymmetric unit is a dimer.

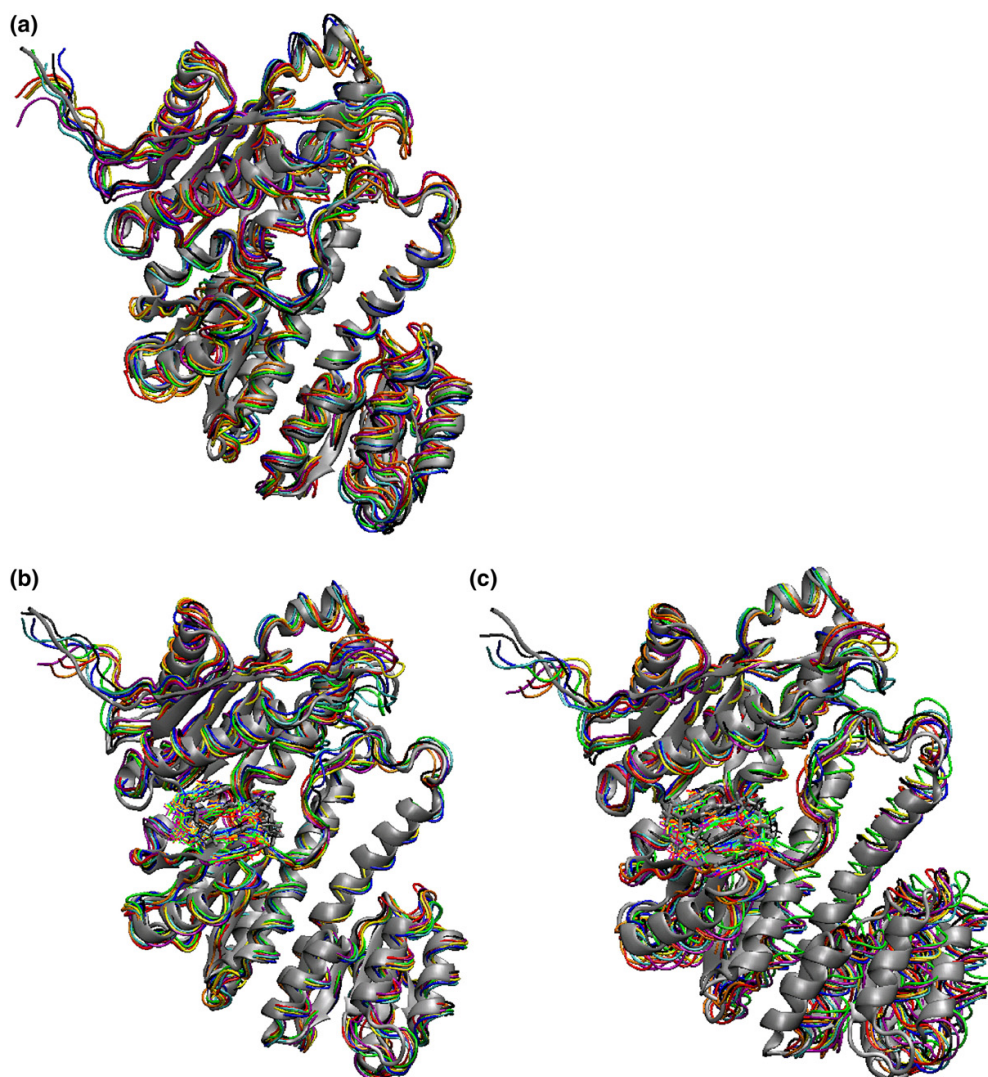
#### Cross-correlation maps for native, ligated PleD

Dynamical cross-correlation maps (DCCMs) allow analysis of persistent (anti)correlated motions of residue pairs from the MD simulations and serve as a quantitative measure for the collective motion in a protein. First, the results obtained for the ligated system are discussed in detail (see Fig. 8a). Then, changes in the DCCMs upon removing the ligand are described. This allows the investigation of ligand-induced effects upon binding in PleD. Correlation coefficients above 0.3 are reported. A threshold of 0.3 is used to capture correlations between the

ligand and the protein. To further characterize the computed correlations, they were calculated for the first and second nanosecond of trajectory I and for shorter (25 ps) and longer (70 ps) windows. The results, shown in Figs. 1 and 2 of the Supplementary Data, demonstrate that the reported correlations are persistent and largely independent of the details of the averaging procedures. A large number of correlated motions within each domain and some interdomain correlations are found, whereas anti-correlated motions are rare. Most correlated motions correspond to spatial proximity of the residues involved.

The correlation pattern of domains D1 and D2 is very similar, as can be expected, because both domains are annotated CheY-like receiver domains. The central  $\beta$ -strand  $\beta 3$  is coupled to all secondary-structure elements of the domain, suggesting a tightly packed domain. For the catalytically active domain (DGC), no single secondary-structure element with correlation to all other elements within DGC are found. Interdomain couplings can serve to “transfer information” (e.g., following a conformational transition in one domain) between individual domains. Such interdomain couplings are cross-correlation motions between different domains. For PleD they are found between neighboring domains, D1/D2 and D2/DGC, respectively. The extended  $\alpha 5$ -helix of D2 is interacting with  $\alpha 4$ ,  $\beta 5$  and  $\alpha 5$  of D1.  $\beta 2$  and  $\alpha 2$  of D2 are correlated to  $\alpha 0$  of DGC.

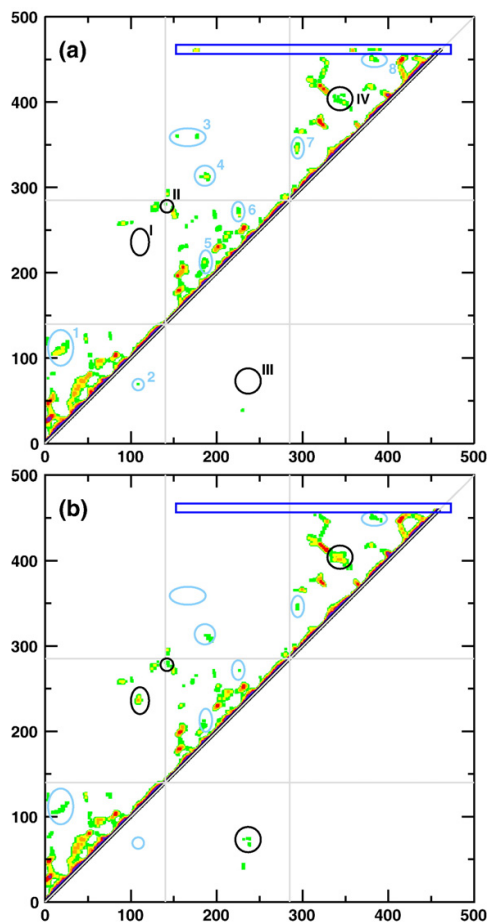
For the loop regions, mainly intradomain correlations are found. The couplings in domains D1 and D2 are again similar. Interdomain correlations are found between D2 and DGC, but none between D1 and D2. The two unstructured regions connecting neighboring domains, D1–D2 and D2–DGC, respectively, are correlated. Additional connections are from  $L(\alpha 2, \beta 2)$ —the I-site loop—to the unstructured region D1–D2 and to  $L(\alpha 1, \beta 2)$ .



**Fig. 7.** Superposition of snapshots for ligated and unligated PleD. Superposition of eight snapshots taken every 250 ps from the MD trajectories on the X-ray structure. The snapshots are reoriented on all backbone atoms for unligated (a) and ligated (b) PleD. Reorientation on the backbone atoms of DGC only for ligated PleD (c). The structures are colored as follows: X-ray structure, silver (cartoon representation); 0.25 ns, black; 0.50 ns, blue; 0.75 ns, cyan; 1.00 ns, green; 1.25 ns, yellow; 1.50 ns, orange; 1.75 ns, red; and 2.00 ns, purple. Comparison of (b) and (c) shows the displacement of D1 and D2 versus DGC and the flexibility of the region  $\beta 3'/\beta 3''$  of DGC.

Removing the ligand from the I-site modifies the cross-correlation pattern. In the following, only changes upon ligand withdrawal are described, first for secondary-structure elements and then for the loop regions. If the ligand is removed from the I-site, changes occur in all three domains. Some correlations present in the complex disappear and new correlations emerge.  $\alpha 4$  of D2 becomes more strongly correlated to elements within D2 and also couples to  $\alpha 5$  of D1, which is now correlated to  $\alpha 2$  of D1. Within DGC, coupling between  $\alpha 1$  and  $\beta 3'$  is

found. It should be noted that these secondary-structure elements are close to the active site. Within DGC, correlation between  $\alpha 0$  and  $\beta 3$  is deleted. This might be of interest, as  $\beta 3$  is part of the  $\beta$ -sheet connecting the I-site with the A-site loop and  $\alpha 0$  is at the interface to domain D2. In D2, interactions between  $\alpha 1$  and  $\alpha 3$  and between  $\beta 1$  and  $\alpha 5$  disappear. In D1,  $\alpha 1$  and  $\beta 3$  are no longer correlated. Residue D53, which is phosphorylated to activate the DGC, is part of  $\beta 3$ . Finally, the interaction between D2 and DGC, namely,  $\beta 2$  of D2 and  $\alpha 0$  of



**Fig. 8.** DCCM for ligated and unligated PleD. DCCMs for ligated (a) and unligated (b) PleD. The following color code is used for the cross-correlation values: 0.3–0.4, green; 0.4–0.5, yellow; 0.5–0.6, orange; 0.6–0.7, red; 0.7–0.8, indigo; 0.8–0.9, blue; 0.9–1.0, black. The blue box marks direct correlations with the ligand. Blue and black circles, labeled with arabic and roman numerals, mark correlations that are present exclusively in ligated and unligated PleD, respectively.

DGC, is weakened. Analysis of the loop regions shows changes concerning all three domains as found for the secondary-structure element analysis.

In D2,  $L(\beta_3, \alpha_3)$  and  $L(\alpha_4, \beta_5)$  become correlated when the ligand is removed. For DGC, a new coupling is found between  $L(\alpha_0, \beta_1)$  and the C-terminus. In addition,  $L(\beta_5, \alpha_5)$  of D1 interacts with  $L(\beta_4, \alpha_4)$  across the D1/D2 interface, whereas interdomain correlation between D2 and DGC is decreased. The I-site loop  $L(\alpha_2, \beta_2)$  no longer interacts with the disordered region D1–D2 and with the loop  $L(\alpha_1, \beta_2)$ . The absence of these interdomain correlations can be rationalized by spatial contacts *via* the ligand.

In summary, ligand removal leads to stronger correlation between D1 and the proximal part of D2. However, ligand-induced couplings, whether due to spatial proximity or to structural rearrangement, disappear and are mainly located at the I-site pocket, the interface D2–DGC, in domain DGC close to the I-site loop,  $\beta_3$ , and in the distant domain D1 at the activation site, D53 of  $\beta_3$ . These trends are observed to be persistent for different trajectories of ligated PleD (see Supplementary Data Fig. 3 for validation and Fig. 4 for a comparison of the DCCMs).

Finally, the direct correlations between PleD and c-di-GMP are discussed. The ligand is coupled to residues from both domains D2 and DGC. The secondary-structure elements  $\alpha_1$  and  $\beta_2$  of D2 and  $\alpha_2$  and  $\alpha_3$  of DGC are correlated with the ligand. The two elements of D2 show different correlations in the ligated and unligated protein, whereas the elements of DGC are unaffected by ligand binding. Furthermore, the ligand is interacting with the following loop regions:  $L(\alpha_1, \beta_2)$  and  $L(\beta_2, \alpha_2)$  of D2 and  $L(\alpha_2, \beta_2)$  (I-site) and  $L(\beta_3, \alpha_3)$  of DGC. The loops  $L(\alpha_1, \beta_2)$  of D2 and  $L(\alpha_2, \beta_2)$  of DGC show differences in loop–loop correlations upon ligand removal.

The correlation patterns for ligated and unligated PleD and the differences between them may help to rationalize how ligand binding at the I-site modifies and/or controls A-site functionality. One possibility is that control occurs directly *via* the short  $\beta$ -strand  $\beta_2$  that directly connects the I-site and the A-site. However, no direct correlation to the ligand is found from the cross-correlation analysis. Another route is coupling through the neighboring  $\beta$ -strand  $\beta_3$  or the  $\alpha$ -helices  $\alpha_2$ ,  $\alpha_3$  situated above the  $\beta$ -sheet. The cross-correlations display couplings from  $\beta_2$  to both the antiparallel  $\beta$ -strand  $\beta_3$  and the helix  $\alpha_0$  situated at the interface to the neighboring domain D2, and the correlation pattern of the neighboring  $\beta$ -strand  $\beta_3$  is sensitive to the presence or absence of the ligand. The motion of the I-site loop is correlated with  $L(\beta_3, \alpha_3)$ , and if the ligand is present the I-site loop is coupled to  $L(\alpha_1, \beta_2)$  of D2 and the disordered region D1–D2. The A-site loop, consisting of the residues important for substrate binding and catalysis, is coupled to the disordered region D2–DGC. From the cross-correlations, a model involving  $\alpha_2$  or  $\alpha_3$  as connection element is less likely since ligand binding does not affect the dynamical correlation of these helices.

Two possible pathways remain: either *via* the  $\beta$ -sheet or *via* the disordered regions. Pathway I involving  $\beta_2$  and  $\beta_3$  is plausible because the secondary-structure elements connect the two sites directly and  $\beta_3$  correlations are altered by ligand binding. This finding is also in accord with the previously reported  $\beta_2$  balance-like movement of the I- and A-site loops, based on the superposition of the minimized structures.<sup>2</sup> Pathway II *via* the disordered regions, D1–D2 and D2–DGC, originates from an observed cascade of direct correlations from the I-site to the A-site in the DCCM analysis. In addition, this cascade is only present if the ligand is



bound, otherwise the connection I-site loop to disordered region D1–D2 is absent.

## Discussion and Conclusions

Currently, it is believed that both structure and dynamics contribute to protein function.<sup>22–26</sup> Complementary to experimental work such as point mutation studies or binding affinity measurements, computer simulations provide valuable and testable information at an atomic level. In this work, the ligated and unligated multidomain DGC PleD monomer is characterized by using MD simulations. Studying the monomer is appropriate because it has been previously proposed that the monomer is crucial for understanding allosteric inhibition in PleD.<sup>8</sup> This is further supported by investigations of the single-domain DGC DgcA, which has preserved the ability for feedback inhibition that indicates that allostery is an intrinsic characteristic of the DGC domain.<sup>2</sup> Finally, binding of c-di-GMP to the PleD monomer and dimer has been proposed to eliminate catalytic activity by domain immobilization.<sup>8</sup> This is in agreement with the observation from the present simulations that ligation of c-di-GMP leads to a rigidification of the monomer.

In the present work, the applicability of atomistic simulations was first validated by calculating optimized structures and RMSFs around the minimum that can be directly compared to experimental data (X-ray structure and B-factors). Having established that this approach is meaningful, we investigated the normal modes and the molecular dynamics in detail. The results suggest that the inhibition (I-site) and the active site (A-site) of PleD are dynamically coupled through an elaborate network of atomic motions that involves the  $\beta$ 2-strand (direct coupling within DGC) and the structural elements  $\beta$ 3,  $\alpha$ 2 and  $\alpha$ 3 of DGC (more delocalized, network coupling).

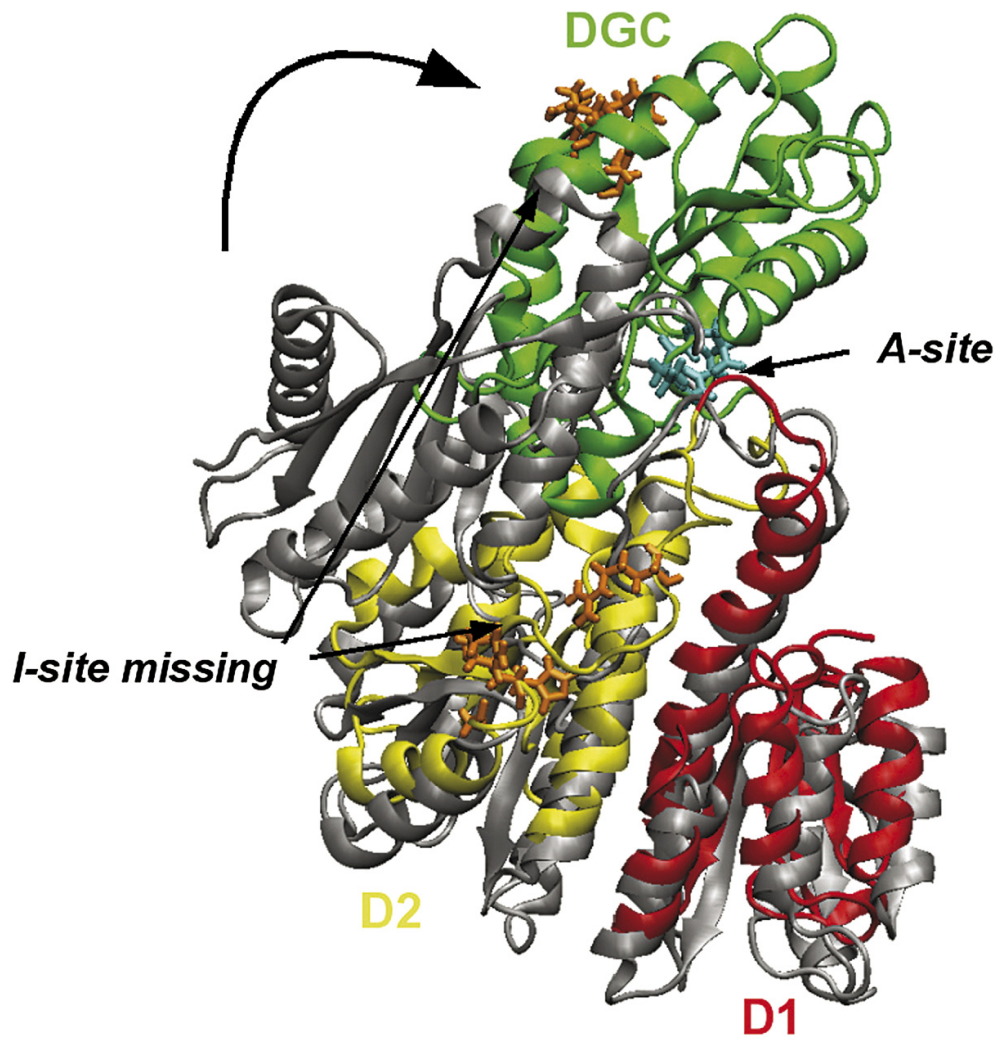
From the superposition of the energy-minimized structures of ligated and unligated PleD a balance-like movement of the  $\beta$ -strand  $\beta$ 2 was previously proposed.<sup>2</sup> This motion displaces the I- and A-site loops simultaneously. The present work extends this model explicitly to the atomic dynamics and does not assume that the ligated (experimentally observed) and the unligated protein structure are closely related, which is not necessarily the case (see below). Thus, the results presented here are stronger evidence that direct communication between I- and A-site along  $\beta$ 2 contributes to allosteric control in PleD.

The experimentally determined structure for PleD was obtained through co-crystallization with c-di-GMP.<sup>8</sup> For further understanding of the mechanisms underlying PleD function additional insight into the unligated structure is useful. The prediction of the structural changes upon ligand binding is a difficult task. Using the protocol described in Materials and Methods [steepest descent (SD)/adopted Newton–Raphson (ABNR) minimization until RMS gradient

of  $10^{-7}$ , distance-dependent dielectric with  $\epsilon=4$ , decreasing constraints on the backbone atoms], snapshots along the MD trajectory were minimized. Assuming a large-scale motion of the DGC domain with respect to domains D1 and D2, an “open” structure (see Fig. 9) that is considerably different from the ligand-bound structure can be envisaged. Ten snapshots from a short, high-temperature MD simulation (10 ps,  $T=500$  K, distance-dependent dielectric with  $\epsilon=4$ ) were minimized using the standard protocol. The final structures are considerably lower in energy (by 95 kcal/mol) compared to optimized structures of unligated PleD in its X-ray conformation. Favorable electrostatic interactions and orientations (dihedral angles) stabilize the putative open structure by  $\approx 140$  kcal/mol relative to the optimized unligated structure of PleD, whereas unfavorable van der Waals and valence angle energies destabilize it by  $\approx 45$  kcal/mol. It is interesting to note that the low-energy character of this postulated open conformation is in qualitative agreement with X-ray data of BeF<sup>-</sup>-activated PleD structure, although the experimentally found structure is even more open.<sup>27</sup> The open, unligated structure has no I-site (originally formed by residues from the DGC and D2 domains), and the A-site is close to the unstructured region connecting domains D1 and D2. The existence of two distinct conformations (unligated open *versus* ligated closed) strengthens the widely known equilibrium shift between active and inactive structure through binding of the ligand.<sup>9,28–31</sup>

From the normal mode analysis, no single functional mode with exceptionally high involvement coefficient was found for the eigenvectors calculated from the unligated or for the ligated structure (see Fig. 10). This is in agreement with recent propositions that single functional modes are characteristic for motor proteins, but not for signaling proteins.<sup>15</sup> In fact, PleD is a signaling protein activated by a stimulus (phosphorylation at D53 in domain D1). However, although normal modes can be helpful in analyzing functionally important motions in a protein, there are limitations in their interpretation, as has been discussed recently.<sup>32</sup> The present findings show that binding of c-di-GMP to the I-site does not affect only one dominant, functionally relevant mode, but rather influences a network of motional degrees of freedom that are relevant for PleD function. From the DCCM, two distinct pathways are plausible: the first runs along the secondary-structure elements  $\beta$ 2 and  $\beta$ 3 with the possibility to include the  $\alpha$ -helices  $\alpha$ 2 and  $\alpha$ 3 situated above the  $\beta$ -sheet, whereas the second pathway includes the disordered regions connecting the different domains, D1–D2 and D2–DGC. The second pathway is only possible if the ligand is bound, which implies that the ligand can act as a switch.

A central question in understanding the function of PleD is to relate its primary function (production of c-di-GMP) with self-inhibition by c-di-GMP. Since the I- and A-sites are both located in the DGC domain, one possibility is that ligation of c-di-GMP

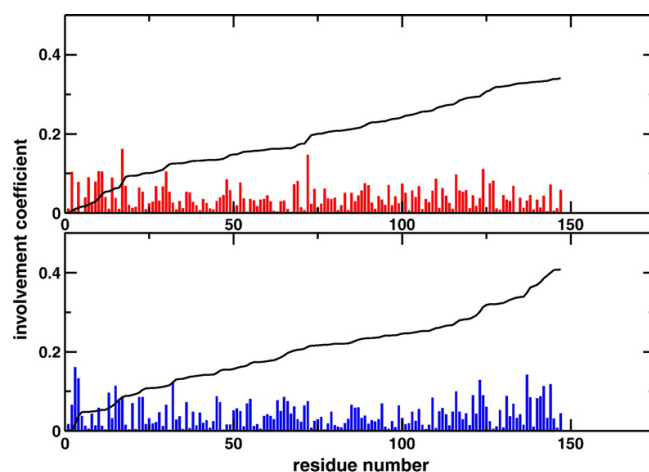


**Fig. 9.** Open structure of free PleD. Putative open conformation structure of PleD superimposed on the original crystal structure (gray): the guanylate cyclase domain (DGC) in green, domains D1 and D2 in red and yellow, respectively. Protein residues forming the active (cyan) and the former inhibition site (orange) present in the crystal structure are highlighted.

in the I-site leads to loss of functional modes (flexibility), which restricts motion around the A-site. Indeed, NMA and RMSF calculations on minimized PleD structures show that c-di-GMP bound at the I-site influences the flexibility of the protein such that specifically both the I- and A-site regions become more rigid. The dynamics simulations support the theory that information transfer between the two functional sites (I- and A-site) might be due to dynamical coupling *via*  $\beta$ -strand  $\beta$ 2 (pathway I) or through delocalized global motion of the protein (pathway II).

To obtain additional information on the proposed pathway and the correlations in inhibited PleD, similar calculations were carried out for the point

mutant R390A with c-di-GMP bound at the I-site.<sup>2</sup> The mutation is located in the  $\alpha$ -helix  $\alpha$ 3 of the DGC domain and was obtained by deleting and rebuilding the appropriate residue side chain. Changes in the correlation pattern are found in all three domains, which indicates that the R390A mutation affects the dynamics of the entire protein. Correlations within domains D1 and D2 are strongly decreased in the mutant compared to the wild-type protein, whereas the correlation between the two domains is increased. In the catalytically active domain DGC a similar number of correlations occur. However, different structural elements are correlated in the native compared to the mutant protein. For pathway I ( $\beta$ 2 and  $\beta$ 3 including the  $\alpha$ -helices  $\alpha$ 2



**Fig. 10.** Involvement coefficients calculated for the transition between minimized structures of ligated and unligated PleD. Involvement coefficients,  $CI_N$ , calculated for the transition between the minimized (free) structure of ligated and unligated PleD from the eigenvectors obtained with DIMB (basis set 153). No single mode contributing most to the structural transition can be found for the eigenvectors calculated from the unligated (top) structure or for those obtained from the ligated (bottom) structure. The cumulative sum of the square of the involvement coefficients is shown as a black line.

and  $\alpha 3$ ) the I- to A-site correlation is largely preserved, whereas the direct correlation of the ligand with  $\alpha 2$  is missing. In contrast, pathway II (connecting regions between domains D1–D2 and D2–DGC) no longer exists. Overall, the DCCM of the mutant R390A shows that a single point mutation considerably influences the overall dynamics in PleD. The mutation alters the correlations across the entire protein, and both suggested pathways for I-/A-site communication are affected. This is in accord with combined mutagenesis and biochemical experiments on the point mutant R390A, which show that the decreased feedback inhibition is mostly due to a strong decrease in c-di-GMP binding affinity.<sup>2</sup>

However, to corroborate their existence and to discriminate between the two pathways, further experimental and computational investigations are required. To verify that the DGC domain is sufficient for feedback inhibition and to exclude pathway II it will be interesting to study the DGC domain of PleD individually by biochemically removing the link to domains D1 and D2. Pathway I involves the motifs  $\beta 2$  and  $\beta 3$ , which are coupled to the  $\alpha$ -helices  $\alpha 2$  and  $\alpha 3$ . Studying the double mutant N357A/H394A (disruption of hydrogen bond between  $\alpha 2$  and  $\alpha 3$ ) or the single mutant R353 (side-chain backbone interaction between  $\alpha 2$  and  $\alpha 3$  will be unavailable) should lead to decreased interaction between the  $\alpha$ -helices that in turn affects coupling between the I- and the A-site.

Based on MD simulations, the present work characterizes the atomic dynamics of ligated and unligated PleD. The mechanistic model proposed by Chan *et al.* is to some extent supported by the present data; for example, an “open” structure of PleD is suggested from preliminary X-ray data and simulation data where the mobility of DGC with respect to domains D1/D2 leads to a structure that is lower in energy.<sup>8</sup> However, ligand binding to the inhibition site leads more likely to the loss of protein flexibility than to a total immobilization of the three

domains. It was found that upon ligand binding the protein is rendered less flexible specifically in both the I- and A-sites, which was shown to be mediated through dynamical coupling between the two sites. Simulations found stronger correlations between D1 and D2 for unligated PleD, which may affect the dimerization rate. Together with the available biological data, this is a first step towards a detailed, atomistic understanding of the important bacterial second messenger c-di-GMP. Further experimental and computational work on this rich and interesting system should provide fundamental information about allosteric control and the role of dynamics in protein–ligand systems.

## Materials and Methods

### Setup of the protein and the ligand

All MD simulations were carried out with either CHARMM<sup>33</sup> or NAMD<sup>34</sup> using the CHARMM 22/27 force field.<sup>35</sup> The starting structure was taken from the X-ray structure of the PleD dimer [Brookhaven Protein Data Bank (PDB), code 1W25] resolved at 2.7 Å. For all subsequent studies the A chain of PleD with the c-di-GMP molecule in the inhibition (I-) site was considered. The additional c-di-GMP molecule in the active (A-) site and the metal ions are omitted because the present study is primarily concerned with investigating effects upon inhibition of PleD (the  $Mg^{2+}$  ion close to residue D53 is important for the phosphorylation reaction). The protein PleD consists of 460 amino acids and a total of 7140 atoms. Hydrogen atoms and the missing residues (M1, His-tag H456–H460) were generated and the structure was relaxed by 500 steps of SD minimization. All titratable side chains are generated in their standard protonation state for pH 7. The system has an overall charge of  $-7$  for PleD and  $-4$  for the ligand. No counterions were introduced to neutralize the systems.

To describe the nonstandard residue c-di-GMP, parameters and partial charges were adopted from the extended CHARMM parameter sets for nucleic acids. The structure of the intercalated c-di-GMP was optimized

using 5000 steps of SD minimization followed by ABNR minimization until a RMS gradient of  $10^{-7}$  was reached in a distance-dependent dielectric with  $\epsilon=4$ . Also, the ligand was solvated in a small ( $37.3 \text{ \AA} \times 34.1 \text{ \AA} \times 34.1 \text{ \AA}$ ) water box and 24 ps of heating and equilibration and 100 ps production MD simulations were run to assess its stability.

The structure of the ligated (intercalated c-di-GMP docked into the inhibition site of PleD) and the unligated protein was optimized using a distance-dependent dielectric with  $\epsilon=4$  and a cutoff of 12  $\text{\AA}$  for nonbonded interactions. Starting from the X-ray structure, 1000 steps of SD optimization were followed by 7500 steps of ABNR minimization with gradually decreasing constraints (more restrictive on the backbone atoms) to prevent large structural distortions. Initially, the structures were relaxed until a RMS gradient of 0.01 kcal/mol was reached. Such a minimization will be referred to as "constrained." Starting from the constraint-optimized structure, further minimization using no constraints and a RMS gradient of  $10^{-7}$  kcal/mol as a convergence criterion will be labeled "constrained E-7." In addition, full minimizations until a RMS gradient of  $10^{-7}$  kcal/mol without any constraints were carried out, which are called "free."

### Normal-mode calculations

Normal mode calculations were carried out with the VIBRAN module of CHARMM. Since the unligated and the ligated protein consist of 7140 and 7276 atoms, respectively, no full NMA is possible because of the large size of the resulting Hessian matrix. Thus, the approximate schemes, DIMB and coarse-grained one-residue-per-block method (BNM) as implemented in CHARMM were used.<sup>21,36–38</sup> The normal modes were calculated for minimized structures of the system. The minimization was performed using SD and ABNR minimization until the RMS of the energy gradient reached a value of 0.01 kcal/mol/ $\text{\AA}$  using gradually decreasing constraints or  $10^{-7}$  kcal/mol/ $\text{\AA}$ . A gradient value of 0.01 kcal/mol/ $\text{\AA}$  has been shown to be satisfactory for calculating normal modes. RMSD from the X-ray structure calculated for the backbone atoms (residues 2–454) is 1.77  $\text{\AA}$  (free) and 0.97  $\text{\AA}$  (constrained) for the complex and 1.88  $\text{\AA}$  (free) and 0.98  $\text{\AA}$  (constrained) for the unligated protein. The basis set size was varied from 50 to 300 to analyze its convergence. The translational–rotational coupling was removed by shifting these modes to high frequency. To test the DIMB approach for the present case, normal modes for the intercalated c-di-GMP ligand were calculated with diagonalization of the full Hessian (DIAG command) and compared to DIMB calculations retaining 103 basis functions.

The BNMs are coarse-grained descriptions that allow one to calculate normal modes at reduced computational expense. BNM uses a coarse-graining procedure that can range from a single residue to a secondary-structure element. For BNM, the recent implementation into the CHARMM program was used.<sup>38</sup> One residue per block was chosen and the frequencies were adjusted by scaling with 0.5882 as suggested.

### Molecular dynamics simulations

For the MD simulations the protein was solvated in a rectangular box. With CHARMM, a  $99.3 \text{ \AA} \times 74.5 \text{ \AA} \times 96.2 \text{ \AA}$  box of pre-equilibrated TIP3P water<sup>39</sup> was used and solvent molecules overlapping with the protein were removed, which gives a system size of  $\approx 70,640$  atoms:

7140 atoms from native protein PleD, 136 atoms from the ligand (c-di-GMP dimer) and 63,500 solvent atoms approximately. In the case of NAMD, the solvation function with a protein–box distance of 15  $\text{\AA}$  was used, resulting in a box with dimensions of  $74 \text{ \AA} \times 112 \text{ \AA} \times 95 \text{ \AA}$ . This system size contains approximately 75,000 atoms. In all simulations, periodic boundary conditions were used, images were updated every 10 time steps and a 12  $\text{\AA}$  cutoff was applied to the shifted electrostatic and switched van der Waals interactions. All MD simulations were carried out using SHAKE<sup>40</sup> to constrain hydrogen atoms and with a time step of 1 fs. For simulations with CHARMM the solvent was equilibrated at 300 K during 29 ps (4 ps solvent heating and 25 ps solvent equilibration) with the protein kept fixed. The entire system was then heated to 300 K and equilibrated for 275 ps before starting the production runs. Total simulation times were 3.0 and 2.0 ns for ligated and unligated PleD, respectively. In the following, trajectory (I) is the CHARMM simulation for ligated PleD. Additional trajectories (II and III, discussed in the Supplementary Data) for ligated PleD, carried out with NAMD, were run for 2.5 ns (started from the X-ray structure and previously equilibrated for 0.5 ns) and 3.5 ns (started from a snapshot at 0.6 ns of the equilibrated system), respectively.

### Analysis of the results

The results of the NMA and the MD simulations were analyzed by calculating involvement coefficients and cross-correlated motions.

*Involvement coefficients* are extracted from the NMA via

$$I_k = \frac{\vec{X}_1 - \vec{X}_2}{|\vec{X}_1 - \vec{X}_2|} \cdot \vec{L}_n \quad (1)$$

where  $\vec{X}_1 - \vec{X}_2$  is the displacement vector between two conformations and  $\vec{L}_n$  is the eigenvector of the  $n$ -th normal mode. Involvement coefficients  $I_k$  quantify the geometrical similarity between the  $n$ -th normal mode and the conformational transition between the structures considered,  $\vec{X}_1$  and  $\vec{X}_2$ . In addition, the cumulative involvement coefficient  $CI_N$  can be calculated.

$$CI_N = \sum_{n=1}^N I_n^2 \quad (2)$$

For a complete orthonormal set of normal-mode eigenvectors,  $CI_N$  is expected to be 1. Thus,  $I_n^2$  gives the maximal percentile contribution of motion along the  $n$ -th mode to the conformational transition between two structures.

*Cross-correlated motions* were calculated from the MD simulations as averages over blocks of different lengths (25, 50 and 70 ps for the CHARMM run and 25, 50, 100, 200 ps for the first NAMD run) after removing the overall translation and rotation of the system. The cross-correlation coefficient  $C_{ij}$  of residues  $i$  and  $j$  is given by

$$C_{ij} = \frac{\langle \Delta r_i \Delta r_j \rangle}{\sqrt{\langle \Delta r_i^2 \rangle \langle \Delta r_j^2 \rangle}} \quad (3)$$

where  $\Delta r_i$  and  $\Delta r_j$  are the displacements of the backbone atoms from the reference position.  $C_{ij}$  varies between 1 and  $-1$ , where 1 corresponds to movement in the same direction (correlated) and  $-1$  to movement in the opposite

direction (anti-correlated). In a DCCM, motion in the same direction (correlated) can be found in the upper left triangle, while motion in opposite directions (anti-correlated) is in the lower right panel. Typically, correlated motions are much more pronounced than anti-correlated motions.  $\alpha$ -Helices are manifest as a broadening of the diagonal, while  $\beta$ -sheets appear off diagonal (ascending for parallel, and descending for antiparallel  $\beta$ -sheets). It should be noted that DCCMs do not give information about the magnitude or the direction of the motion.

## Acknowledgements

The authors acknowledge financial support from the Swiss National Science Foundation through a Förderungsprofessur to M.M. The authors are grateful for stimulating discussions with Prof. U. Jenal and Prof. T. Schirmer. Generous allocation of computing time for the NAMD simulations at the CSCS in Manno, Switzerland, is acknowledged.

## Supplementary Data

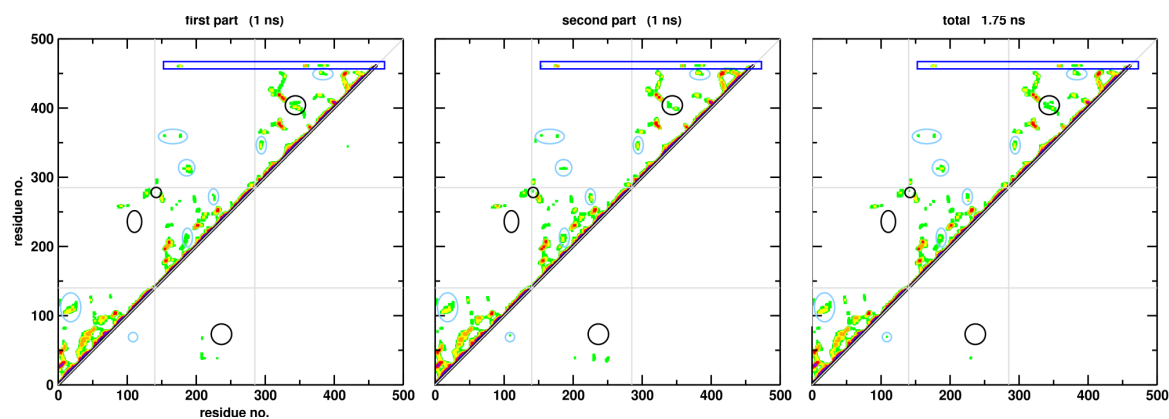
Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.jmb.2007.09.068](https://doi.org/10.1016/j.jmb.2007.09.068)

## References

- Jenal, U. & Malone, J. (2006). Mechanisms of cyclic-di-GMP signaling in bacteria. *Annu. Rev. Genet.* **40**, 385–407.
- Christen, B., Christen, M., Paul, R., Schmid, F., Folcher, M., Jenoe, P. *et al.* (2006). Allosteric control of cyclic di-GMP signaling. *J. Biol. Chem.* **281**, 32015–32024.
- Hisert, K. B., MacCoss, M., Shiloh, M. U., Darwin, K. H., Singh, S., Jones, R. A. *et al.* (2005). A glutamate-alanine-leucine (EAL) domain protein of *Salmonella* controls bacterial survival in mice, antioxidant defence and killing of macrophages: role of cyclic diGMP. *Mol. Microbiol.* **56**, 1234–1245.
- Jenal, U. (2004). Cyclic di-guanosine-monophosphate comes of age: a novel secondary messenger involved in modulating cell surface structures in bacteria? *Curr. Opin. Microbiol.* **7**, 185–191.
- Paul, R., Weiser, S., Amiot, N. C., Chan, C., Schirmer, T., Giese, B. & Jenal, U. (2004). Cell cycle-dependent dynamic localization of a bacterial response regulator with a novel di-guanylate cyclase output domain. *Genes Dev.* **18**, 715–727.
- Hecht, G. B. & Newton, A. (1995). Identification of a novel response regulator required for the swarmer-to-stalked cell transition in *Caulobacter crescentus*. *J. Bacteriol.* **177**, 6223–6229.
- Aldridge, P., Paul, R., Goymer, P., Rainey, P. & Jenal, U. (2003). Role of the GGDEF regulator PleD in polar development of *Caulobacter crescentus*. *Mol. Microbiol.* **47**, 1695–1708.
- Chan, C., Paul, R., Samoray, D., Amiot, N. C., Giese, B., Jenal, U. & Schirmer, T. (2004). Structural basis of activity and allosteric control of diguanylate cyclase. *Proc. Natl. Acad. Sci. USA*, **101**, 17084–17089.
- Formanek, M. S., Ma, L. & Cui, Q. (2006). Reconciling the “old” and “new” views of protein allostery: A molecular simulation study of Chemotaxis Y protein (CheY). *Proteins: Struct. Funct. Bioinf.* **63**, 846–867.
- Lee, S.-Y., Cho, H. S., Pelton, J. G., Yan, D., Berry, E. A. & Wemmer, D. E. (2001). Crystal structure of activated CheY. Comparison with other activated receiver domains. *J. Biol. Chem.* **276**, 16425–16431.
- Zhu, X., Amsler, C. D., Volz, K. & Matsumura, P. (1996). Tyrosine 106 of CheY plays an important role in chemotaxis signal transduction in *Escherichia coli*. *J. Bacteriol.* **178**, 4208–4215.
- Appleby, J. L. & Bourret, R. B. (1998). Proposed signal transduction role for conserved CheY residue Thr87, a member of the response regulator active-site quintet. *J. Bacteriol.* **180**, 3563–3569.
- Volz, K. & Matsumura, P. (1991). Crystal structure of *Escherichia coli* CheY refined at 1.7-Å resolution. *J. Biol. Chem.* **266**, 15511–15519.
- Berendsen, H. J. C. & Hayward, S. (2000). Collective protein dynamics in relation to function. *Curr. Opin. Struct. Biol.* **10**, 165–169.
- Li, G. & Cui, Q. (2004). Analysis of functional motions in “Brownian molecular machines” with an efficient block normal mode approach: Myosin-II and Ca<sup>2+</sup>-ATPase. *Biophys. J.* **86**, 743–763.
- Cui, Q., Li, G., Ma, J. & Karplus, M. (2004). A normal mode analysis of structural plasticity in the biomolecular motor F<sub>1</sub>-ATPase. *J. Mol. Biol.* **340**, 345–372.
- Guex, N. & Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling. *Electrophoresis*, **18**, 2714–2723.
- Levitt, M., Sander, C. & Stern, P. S. (1985). Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme. *J. Mol. Biol.* **181**, 423–447.
- Brooks, B. & Karplus, M. (1983). Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proc. Natl. Acad. Sci. USA*, **80**, 6571–6575.
- Go, N., Noguti, T. & Nishikawa, T. (1983). Dynamics of a small globular protein in terms of low-frequency vibrational modes. *Proc. Natl. Acad. Sci. USA*, **80**, 3696–3700.
- Tirion, M. M. (1996). Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.* **77**, 1905–1908.
- Mayer, K. L., Earley, M. R., Gupta, S., Pichumani, K., Regan, L. & Stone, M. J. (2003). Covariation of backbone motion throughout a small protein domain. *Nat. Struct. Biol.* **10**, 962–965.
- Popovych, N., Sun, S., Ebright, R. H. & Kalodimos, C. G. (2006). Dynamically driven protein allostery. *Nat. Struct. Biol.* **13**, 831–838.
- Vendruscolo, M. & Dobson, C. M. (2006). Structural biology: dynamic visions of enzymatic reactions. *Science*, **313**, 1586–1587.
- Hammes-Schiffer, S. & Benkovic, S. J. (2006). Relating protein motion to catalysis. *Annu. Rev. Biochem.* **75**, 519–541.
- Eisenmesser, E. Z., Bosco, D. A., Akke, M. & Kern, D. (2002). Enzyme dynamics during catalysis. *Science*, **295**, 1520–1523.
- Wassmann, P., Chan, C., Paul, R., Beck, A., Heerklotz, H., Jenal, U. & Schirmer, T. (2007). Structure of BeF<sub>3</sub> modified response regulator PleD: implications for diguanylate cyclase activation, catalysis, and feedback inhibition. *Structure*, **15**, 1–13.

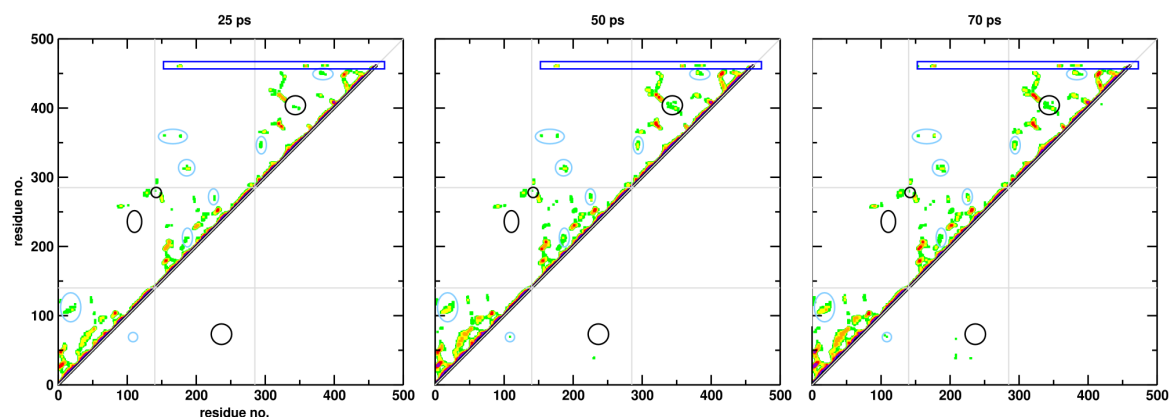
28. Gunasekaran, K., Ma, B. & Nussinov, R. (2004). Is allostery an intrinsic property of all dynamic proteins? *Proteins: Struct. Funct. Bioinf.* **57**, 433–443.
29. Stock, A. M. & Guhaniyogi, J. (2006). A new perspective on response regulator activation. *J. Bacteriol.* **188**, 7328–7330.
30. Volkman, B. F., Lipson, D., Wemmer, D. E. & Kern, D. (2001). Two-state allosteric behavior in a single-domain signaling protein. *Science*, **291**, 2429–2433.
31. Kern, D. & Zuiderweg, E. R. P. (2003). The role of dynamics in allosteric regulation. *Curr. Opin. Struct. Biol.* **13**, 748–757.
32. Wynsberghe, A. W. V. & Cui, Q. (2006). Interpreting correlated motions using normal mode analysis. *Structure*, **14**, 1647–1653.
33. Brooks, B. R., Brucoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983). CHARMM: a program for macromolecular energy, minimization and dynamics calculations. *J. Comput. Chem.* **4**, 187–217.
34. Phillips, J. C., Braun, R., Wang, W., Gumbardt, J., Tajkhorshid, E., Villa, E. *et al.* (2005). Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802.
35. MacKerell, A. D., Jr, Bashford, D., Bellott, M., Dunbrack, R. L., Jr, Evanseck, J. D., Field, M. J. *et al.* (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, **102**, 3586–3616.
36. Mouawad, L. & Perahia, D. (1993). Diagonalization in a mixed basis—a method to compute low-frequency normal-modes for large macromolecules. *Biopolymers*, **33**, 599–611.
37. Tama, F., Gadea, F. X., Marques, O. & Sanejouand, Y. H. (2000). Building-block approach for determining low-frequency normal modes of macromolecules. *Proteins: Struct. Funct. Genet.* **41**, 1–7.
38. Li, G. H. & Cui, Q. (2002). A coarse-grained normal mode approach for macromolecules: an efficient implementation and application to Ca<sup>2+</sup>-ATPase. *Biophys. J.* **83**, 2457–2474.
39. Jorgensen, W. L., Chandrasekhar, J., Madura, J., Impey, R. & Klein, M. (1983). Comparison of simple potential functions for simulating liquid water. *J. Comput. Phys.* **79**, 926–935.
40. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. (1977). Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J. Comput. Phys.* **23**, 327–341.

## Appendix A: Supplementary Data



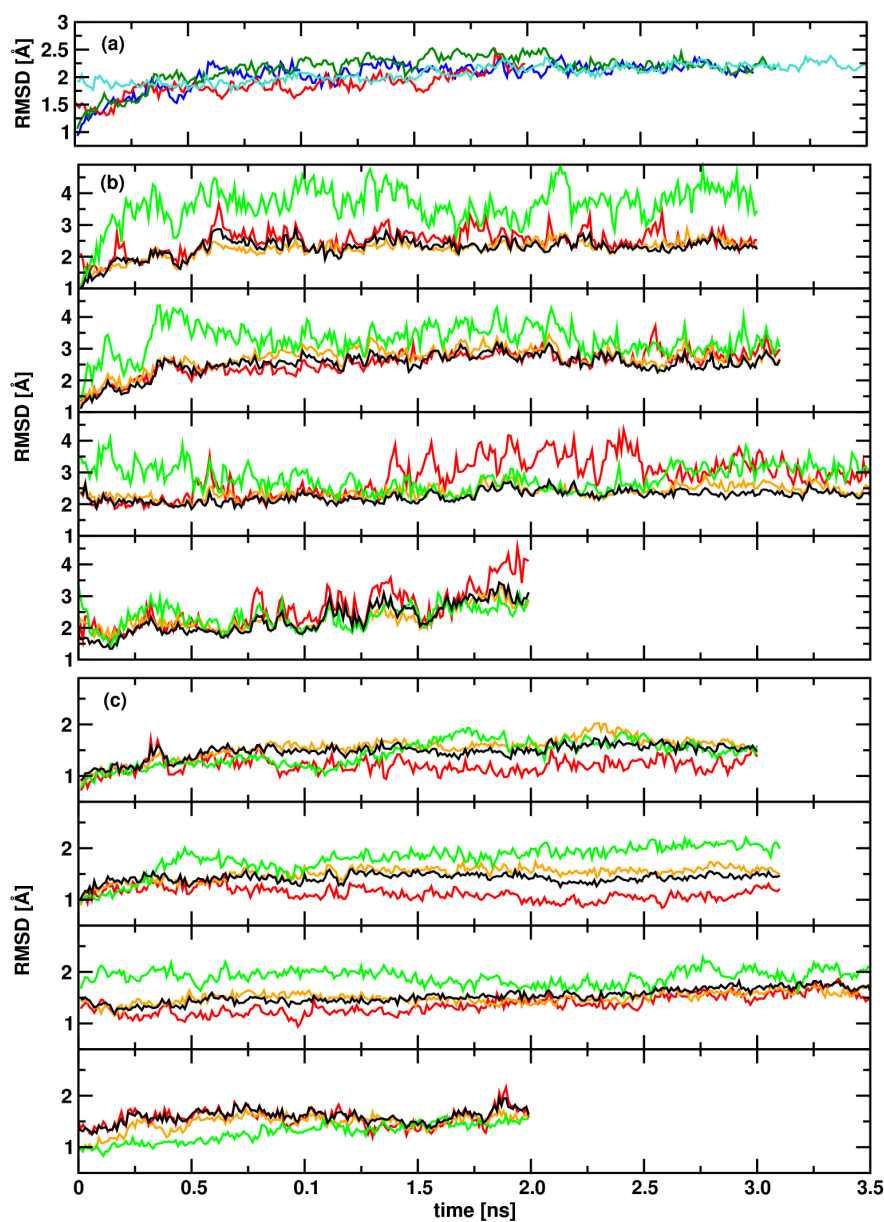
**Fig. 1. DCCMs for first and second half of trajectory I.**

Cross-correlation maps (DCCM) calculated from the first and second half of trajectory I are shown for ligated PleD. The main features are found in all simulation parts. The following color code is used for the cross-correlation values: 0.3-0.4 (green), 0.4-0.5 (yellow), 0.5-0.6 (orange), 0.6-0.7 (red), 0.7-0.8 (indigo), 0.8-0.9 (blue), 0.9-1.0 (black).



**Fig. 2. Influence of window length on DCCMs from trajectory I.**

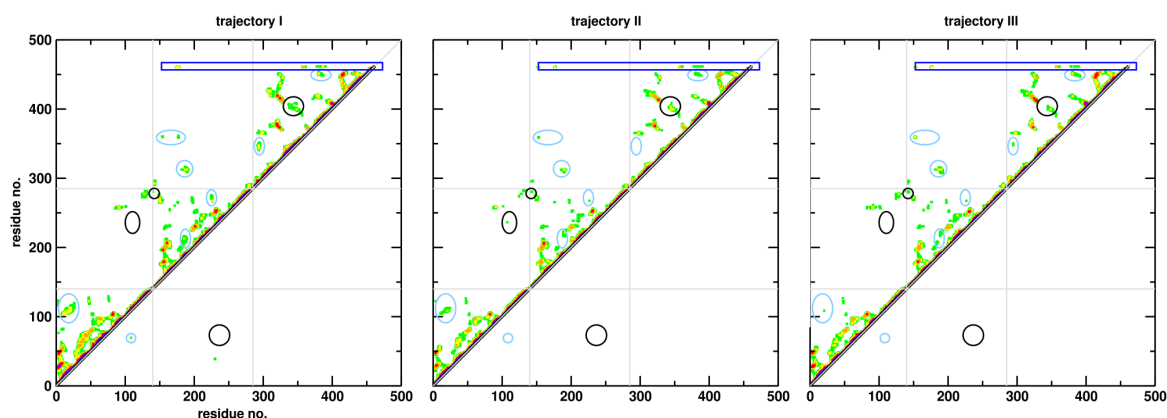
Cross-correlation maps (DCCM) for ligated PleD are shown for different window lengths. The correlation pattern between different structural elements is persisting. The intensity nevertheless may change. Color code as in Figure 1 Supplement.



**Fig. 3. RMSD of additional MD simulations for ligated PleD.**

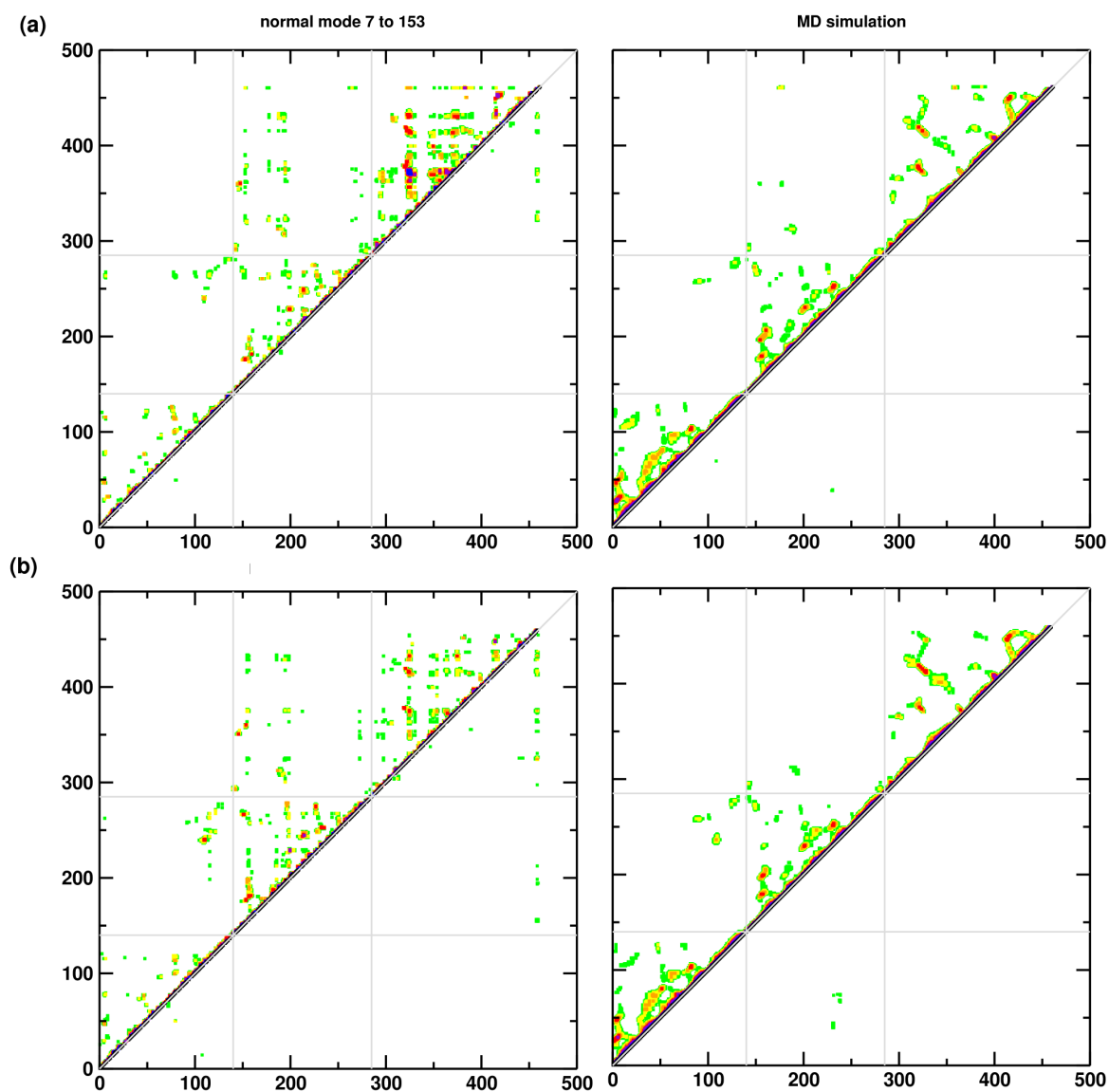
Root mean square displacements (RMSD) for  $C_{\alpha}$ -atoms with respect to the X-ray structure of ligated PleD from additional MD simulations in explicit solvent. (a) the RMSD calculated for all  $C_{\alpha}$ -atoms is  $\approx 2.25$  Å for all trajectories. Ligated PleD trajectory I (blue), trajectory II (dark green), trajectory III (turquoise) and unligated PleD (red). (b) RMSD for all  $C_{\alpha}$ -atoms after superposition on domains D1 (red), D2 (orange), DGC (green) and D1/D2 (black) respectively. From top to bottom: trajectory I, II, III and unligated PleD. (c) RMSD calculated for each domain separately (same color code; same order top to bottom).





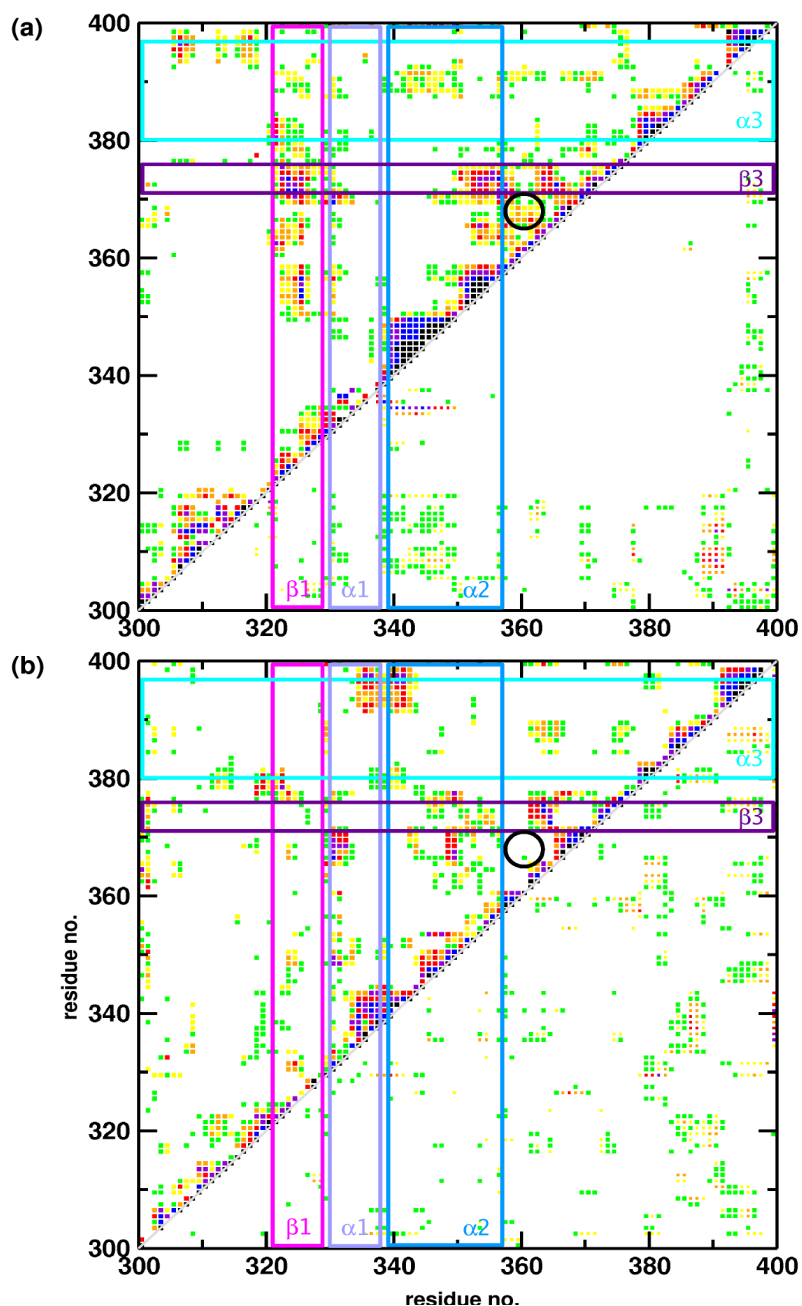
**Fig. 4. DCCM for trajectories I, II and III.**

Cross-correlation maps (DCCM) for ligated PleD are shown for all three trajectories calculated. Differences in correlations, especially in their intensity, are observed. It is found that correlations within domains vary more compared to correlations between domains which are either present or absent in all three simulations respectively (correlations I, III, 3 and 4). The direct correlations of the ligand with protein parts in the D2 and DGC domain are found in all three trajectories. Characteristic features for ligated PleD are the direct interactions with the ligand which induces inter-domain correlations between D2 and DGC (correlations 3 and 4). Particularly the I-site loop is contributing. It should be noted that for the “open” (unligated) structure D2/DGC correlations are absent. Unligated PleD has major correlations between domains D1 and D2 (correlation I) and stronger coupling of  $\alpha 2$  and  $\alpha 3$  of DGC (correlation IV). Labeling of the prominent correlations as in Figure 8 and color code as in Figure 1 Supplementary Material.



**Fig. 5. Normal mode covariance matrix for ligated (a) and unligated (b) PlcD.**

Covariance matrices for normal modes 7 to 153 already contain features consistent with those found in the MD simulations. However, the correlations from the NMA should be interpreted and compared to DCCMs from MD simulations with care.<sup>32</sup> Color code as in Figure 1 Supplement.



**Fig. 6. Covariance matrix for modes 90 to 100 of ligated and unligated PleD.**

Comparison of the DCCMs for residues in the DGC domain from a superposition of normal modes 90 to 100 in ligated (a) and unligated (b) PleD. The most prominent differences between (a) and (b) are the more pronounced coupling of the A- and I-site residues (marked by a circle) and the coupling of the structural elements  $\beta 1$  with  $\alpha 2$  and  $\beta 1$  with  $\alpha 3$  when the ligand c-di-GMP is present. Secondary structure elements in the DGC domain are indicated by boxes. Color code as in Figure 1 Supplement.

## 5.3 PleD dimerization process

### Introduction

Sensing environmental signals and rapid metabolic response are crucial for bacteria to survive. A typical cascade to integrate external stimuli is the so-called "two-component" signal transduction system<sup>162</sup>, i.e. the signal is sensed by a histidine protein kinase that transfers a phosphoryl group to a conserved aspartate of a response regulator (RR) protein thereby modulating its activity. The diguanylate cyclase PleD of *Caulobacter crescentus* is such a response regulator with its activity controlled by the kinases DivJ and PleC and is required for polar differentiation in the cell cycle of the bacteria.<sup>137,147,157,163</sup> Cells without functional PleD are hypermobile and fail to accomplish the swarmer-to-stalked cell transition. PleD synthesizes the bacterial second messenger cyclic di-guanylic monophosphate (c-di-GMP)<sup>147</sup>, a molecule of great interest since it regulates surface-adhesion properties and motility in bacteria<sup>130,141</sup>. In addition, relevance to virulence of pathogenic bacteria and implication in biofilm formation has been demonstrated.<sup>142</sup> The fact that c-di-GMP is found exclusively in bacteria makes it a potential target for medicinal applications.

The symmetric condensation reaction of two GTP molecules into c-di-GMP requires the encounter of two loaded active sites, i.e. exclusively dimeric PleD is catalytically active which has been confirmed by cross-linking experiments.<sup>158</sup> Thus, the overall production rate depends on the dimerization rate. PleD dimerization is significantly enhanced upon phosphorylation (or pseudo-phosphorylation by  $\text{BeF}_3^-$ ).<sup>158</sup> This phosphorylation-mediated dimerization control constitutes an additional level of allosteric regulation. The unit cell in the X-ray structures of both inactive and  $\text{BeF}_3^-$ -activated PleD<sup>51,52</sup> contain dimeric PleD indicating that the stem domains D1/D2 form the dimerization interface.  $\text{BeF}_3^-$ -activated PleD shows an optimal tight-packed dimer interface that comes about by a shift/rotation of domain D1 with respect to D2.<sup>52</sup> The domains D1 and D2 of PleD are CheY-like receiver domains (REC), whereof only D1 contains all invariant residues of the active site<sup>136</sup>, including the phospho-acceptor residue D53. Structural changes following the phosphorylation have been studied extensively on the model system CheY and are also characteristic to receiver domains.<sup>37,50,168,169,171</sup> The rather subtle change involves the repositioning of the  $\beta$ 4- $\alpha$ 4 loop and sidechain reorientations of conserved Thr/Ser and Phe/Tyr residues also known as "Y-T" coupling. Threonine forms a hydrogen bond to the phospho-moiety and makes space for the Phenylalanine to adopt the "buried" position. This raises the

question of how the local phosphorylation-induced loop repositioning is propagated to the  $\alpha 4$ - $\beta 5$ - $\alpha 5$  interface between domains D1 and D2, and in turn promotes domain reorientation and dimerization. In addition, the exact role of the pseudo-receiver domain D2 is unknown. It could either have an adaptive function responding to the activation state of D1 or solely enlarge the dimerization contact area. And last, mutation experiments showed that the isologous contact residue Y26 is required for dimerization.<sup>158</sup> This residue is located at the center of a hydrophobic patch which exists in both the active and inactive state and it is also strictly conserved in PleD homologs that share the REC-REC-DGC domain architecture. However, its exact role in the dimerization process is unidentified. The present computational study focuses on the characterization of processes associated with the spontaneous relaxation from the active to the non-active state.

## Materials & Methods

### Setup of the "PleD" dimers

All MD simulations were carried out with NAMD<sup>172</sup> using the CHARMM22 force field<sup>74</sup>. From the X-ray structures of inactive and  $\text{BeF}_3^-$ -activated PleD (PDB codes 1W25<sup>51</sup> and 2V0N<sup>52</sup>), residues 1 to 285 which correspond to domains D1 and D2 were kept, deleting all other residues, ions and additional ligand molecules. Simulation of domains D1/D2 should be sufficient to investigate the dimerization process including the observed rearrangement of domains D1 and D2 of PleD. The catalytically active domain DGC shows different orientations in the two X-ray structures arguing for substantial orientational freedom of the domain *vs.* the PleD D1/D2 stem. Absence of the stabilizing  $\text{BeF}_3^-$  moiety in the active conformation will possibly lead to spontaneous active-inactive transition events. All titratable sidechains were generated in their standard protonation state for pH 7, the HSD model was applied to all Histidine residues and the overall charge of  $-3$  in each monomer was not neutralized by adding counterions. Positions of added hydrogen atoms were relaxed using 200 steps of steepest descent minimization. Mutations were introduced by deleting and rebuilding the sidechain of the specific amino acid. The systems studied here are the dimers of domains D1/D2 of PleD in the inactive (*dimer*) and  $\text{BeF}_3^-$ -activated state (*adimer*) and the mutants Y26A in the active conformation, *A\_Y26A* and *B\_Y26A* introducing the mutation in monomer A or B, respectively.

### Molecular dynamics simulations

MD simulations were carried out in explicit TIP3 water solvent. The NAMD solvation package was used with a protein-box distance of 15 Å resulting in system sizes of 80,000 and 50,000 atoms approximately for dimeric and monomeric D1/D2, respectively. Periodic boundary conditions were used in all simulations and images were updated every 10 steps. A 9 Å cutoff was used for shifted electrostatic and switched van der Waals interactions, hydrogen atoms were constrained with SHAKE<sup>173</sup> and a timestep of 1 fs was used. Starting from the solvated X-ray structure, the system was minimized for 1,000 steps and simulations were carried out at  $T = 300$  K and constant pressure of 1 atm. 3 ns trajectories were run and analyzed after 5 ps of equilibration.

### Data analysis

The  $\beta 4$ - $\alpha 4$  loop conformation which is related to the activation state of response regulators can be quantified by the pseudodihedral angle formed by the  $C_\alpha$ -atoms of the loop residues (residue 83 to 86 in domain D1, residue 232 to 235 in domain D2).<sup>37,169</sup> An angle of 113° corresponds to the fully active conformation, whereas 14° is found in the inactive state.

*Interface residues* are defined as all residues that have at least one atom within 5 Å distance of the other monomer or domain.

The *dimerization contact area* is calculated by the difference in solvent accessible surface area (SASA) between the individual monomers and the complex. To calculate the SASA a probe radius of 1.4 Å was used.

The *secondary structure* was assigned for snapshots taken every 10 ps using the sequence analyzer<sup>174</sup> implemented in VMD<sup>175</sup>.

*Cross-correlated motions* were calculated from MD simulations as averages over blocks of 50 ps after removing the overall translation and rotation of the system. The cross-correlation coefficient  $C_{ij}$  of residues  $i$  and  $j$  is given by

$$C_{ij} = \frac{\langle \Delta r_i \Delta r_j \rangle}{\sqrt{\langle \Delta r_i^2 \rangle \langle \Delta r_j^2 \rangle}} \quad (5.1)$$

where  $\Delta r_i$  and  $\Delta r_j$  are the displacements of the backbone atoms from the reference position.  $C_{ij}$  varies between 1 for perfect correlation (motion in same direction) and -1 for anti-correlation (motion in opposite direction), but gives no information about the magnitude or the direction of the motion.

## Results & Discussion

The PleD dimerization process is studied using domains D1/D2 which are known to interact. Considering only the D1/D2 stem domains should be legitimate since the DGC domain is attached via a flexible linker and can move substantially as seen from the two X-ray structures.<sup>51,52</sup> This truncated PleD, the stem D1/D2, is henceforth called PleD. Simulations were initiated from both the  $\text{BeF}_3^-$ -activated structure (1V0N) and the inactive one (1W25), named hereafter *adimer* and *dimer*, respectively. Because force field parameters for  $\text{BeF}_3^-$  or aspartyl-phosphate are not available the  $\text{BeF}_3^-$  moiety was removed from the active conformation before starting the simulations. Instead of studying the active state, the relaxation from the active to the inactive form is investigated here and events accompanying this transition are reported. Moreover, the contact residue Y26 is being mutated to an Alanine in one of the monomer to study its effect on the dimerization. These simulations start from the active conformation and are denoted *A\_Y26A* and *B\_Y26A* depending in which monomer the mutation is introduced.

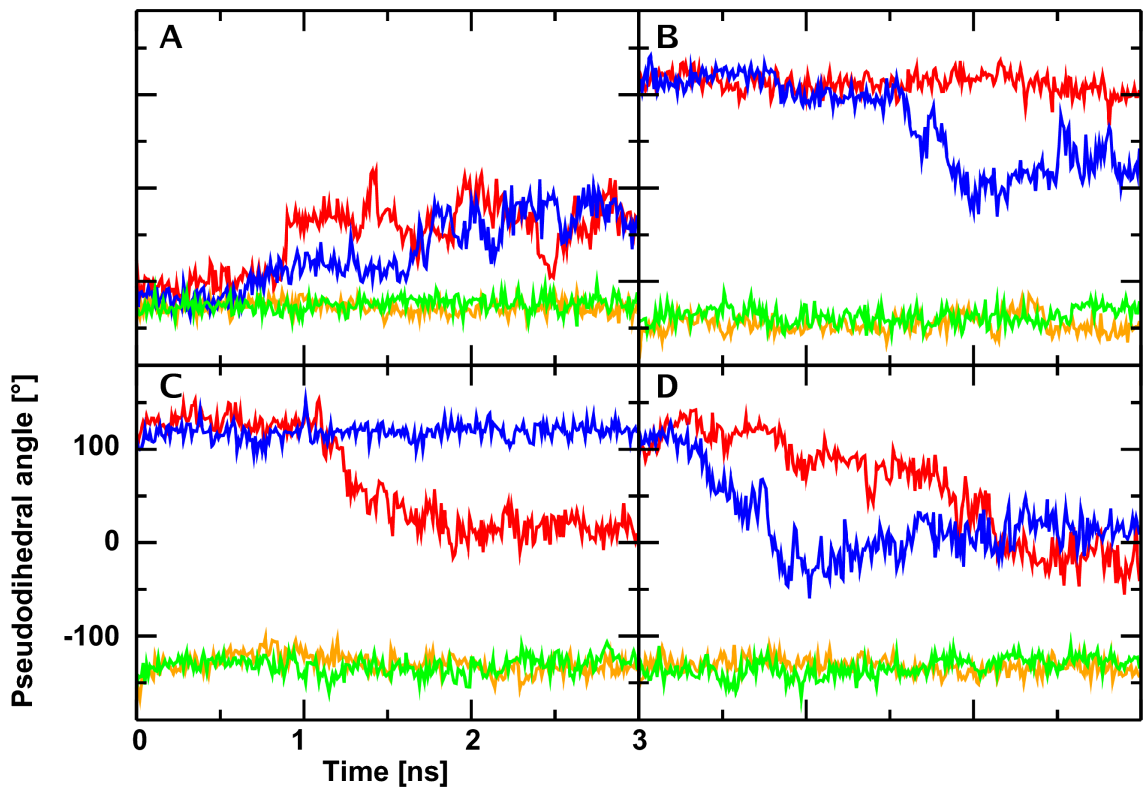
### Structural changes characteristic to receiver domains

#### Spontaneous relaxation of the $\beta 4$ - $\alpha 4$ loop

Domains D1 and D2 of PleD are both assigned to the class of CheY-like receiver domains. These regulatory domains undergo a repositioning of the  $\beta 4$ - $\alpha 4$  loop which is accompanied by a sidechain rotation of the residues Threonine and Tyrosine, a process often termed "Y-T" coupling. The conformation of the  $\beta 4$ - $\alpha 4$  loop can be quantified by the pseudodihedral angle of the loop's  $C_\alpha$ -atoms<sup>37,169</sup>, residues 83 to 86 in domain D1 and residues 232 to 235 in D2 according to the alignment by Hecht et al.<sup>136</sup>. X-ray structures showed that an angle of  $113^\circ$  corresponds to the fully active conformation, while  $14^\circ$  is found in the inactive state.

The pseudodihedral angle values of the  $\beta 4$ - $\alpha 4$  loop sampled during the 3 ns simulation are shown in Figure 7.

*Active-to-inactive transition in domain D1.* In the X-ray structure of  $\text{BeF}_3^-$ -activated PleD an angle of  $100^\circ$  approximately is found which is consistent with findings in CheY and confirms the active conformation. In all three simulations starting from the active form with the  $\text{BeF}_3^-$  moiety removed, a spontaneous active-to-inactive transition of the  $\beta 4$ - $\alpha 4$  loop is observed in at least one of the monomers (see red and blue lines

Figure 7:  $\beta 4$ - $\alpha 4$  loop conformation

The pseudodihedral angle of the  $\beta 4$ - $\alpha 4$  loop in domains D1 and D2 is shown for simulations (A) *dimer*, (B) *adimer*, (C) *A\_Y26A* and (D) *B\_Y26A*. Loop relaxation from the active to the inactive state, indicated by a drop of the pseudodihedral angle value from  $113^\circ$  to  $14^\circ$ , is observed in domain D1 during the 3 ns trajectory, while in D2 no change is observed. Color coding is as follows: red and yellow for domains D1 and D2 in monomer A, blue and green for D1 and D2 in monomer B.

in Figure 7 B, C and D). In *adimer* the loop repositioning is observed at 1.8 ns for monomer B. In the Y26A mutants the first loop transition occurs earlier compared to wildtype PleD, namely at 1.1 and 0.3 ns in simulations *A\_Y26A* and *B\_Y26A*, respectively. The relaxation process is fairly fast and is completed within 0.5 ns approximately. In contrast, the deactivation of the second monomer in *B\_Y26A* is much slower and takes 1.3 ns to complete. Interestingly, this transition closely follows the fast deactivation in the other monomer. On the other hand, in the inactive crystal structure an angle of  $-91^\circ$  is found which is significantly different from the  $14^\circ$  observed in inactive CheY, the model system of RRs. It is possible that this deviation is due to crystal contacts. Moreover, during the *dimer* simulation the pseudodihedral angle changes from  $-91^\circ$  to  $-20^\circ$  approximately (see Figure 7 A), which is closer to the commonly found  $14^\circ$ . Although the value  $14^\circ$  is not reached during the 3 ns simula-



tion time, it indicates that the  $\beta 4$ - $\alpha 4$  loop position found in the X-ray structure is an artefact and an angle between  $0^\circ$  and  $-25^\circ$  is more representative for the inactive state of PleD.

*The dimer stabilizes the active state conformation.* In addition to dynamics simulations of PleD dimer, monomeric PleD was studied to elucidate the influence of an associated monomer. In monomer A of the wildtype (wt) and the Y26A mutant the active-to-inactive transition occurs at 1.2 ns, which is the same time as in the *A\_Y26A* simulation (results for monomer simulations not shown), but does not complete entirely and remains at an angle of  $50^\circ$  approximately. In contrast, for monomer B differences between the wt and mutant are observed. In the former the loop reorientation takes place at 0.2 ns which is much earlier than the 1.6 ns in the dimer simulation. On the contrary, the loop conformation does not change for the Y26A mutant, at least not within the 3 ns calculated. In the case of the inactive monomer the  $\beta 4$ - $\alpha 4$  loop stays significantly longer in the unusual conformation found in the X-ray structure and the pseudodihedral angle changes to a value close to  $0^\circ$  only for monomer B after 2.4 ns. Altogether in PleD wt the loop relaxation occurs earlier in the monomer than in the dimer arguing that an associated monomer slows down the deactivation process by stabilizing the active conformation. This suggestion is in accord with the fact that phosphorylation increases the dimeric fraction. Moreover, the fact that the pseudodihedral angle in the inactive form ranges from  $-90^\circ$  to  $+100^\circ$ ,  $-90^\circ$  for inactive monomer,  $0^\circ$  for inactive dimer and  $+100^\circ$  in the active state, potentially allows "stepwise" regulation.

*Domain D2 shows no change in the loop conformation.* Since domain D2 is also annotated CheY-like domain, we monitored the  $\beta 4$ - $\alpha 4$  loop conformation (see yellow and green lines in Figure 7). D2 shares the common  $(\beta\alpha)_5$  domain fold, though it lacks several conserved residues constituting the active site, including the phospho-acceptor aspartate. In all simulations, dimeric and monomeric PleD in the different activation states, the pseudodihedral angle of the  $\beta 4$ - $\alpha 4$  loop stays close to  $-130^\circ$  the value found in both X-ray structures. Thus, it seems that the  $\beta 4$ - $\alpha 4$  loop positioning in domain D2 is invariant and not correlated to the activation state of PleD. The functional role of D2 is still unknown, yet it seems that domain D2 is not responding to the activation state of D1 in a REC-like fashion, i.e. changing the loop conformation. It is more likely that activation of D1 alters the domain interaction surface and enables domain rearrangement eventually enlarging the dimerization area.

### "Y-T" coupling

The concurrent movement of residues Tyr/Phe and Thr/Ser (F102 and T83 in PleD) is a key signaling cascade in CheY known as "Y-T" coupling. Phosphorylation transfer is followed by the sidechain reorientation of Tyr/Phe from a solvent exposed ( $g^+$ ) to a "buried" position (*trans*). Simultaneously Thr/Ser rearranges to form a hydrogen bond to the phospho moiety. The sidechain orientation can be characterized by the dihedral angle  $\chi_1$ . Thus, we monitored the residue's  $\chi_1$  angle in order to detect any reorientation. Specifically the rotameric state of residues F102 and T83 were analyzed and in general no reorientation is observed during the 3 ns simulations. Only in monomer A of the *active* trajectory a transition from *trans* indicating the active form to *gauche*<sup>-</sup> is observed at 2.5 ns. It is possible that the sidechain rotation of F102 may be observed at longer time scales. However, in PleD residue F102 is located in the D1/D2 interface, while in CheY it is positioned at the surface. This may hinder rotation and in addition there's no "protein-solvent" change between the two rotamers. The counterparts in domain D2, residues I251 and V232, are more flexible. As domain D2 is likely to have a function different from REC domains, there is no relation between their rotameric state and protein activity. Hence, the rotamer sampling is less restricted than for residues involved in the signaling cascade. All characteristics regarding the state of CheY-like RRs are summarized in Table 6 together with events observed in the dynamics simulations.

### D1/D2 domain reorientation

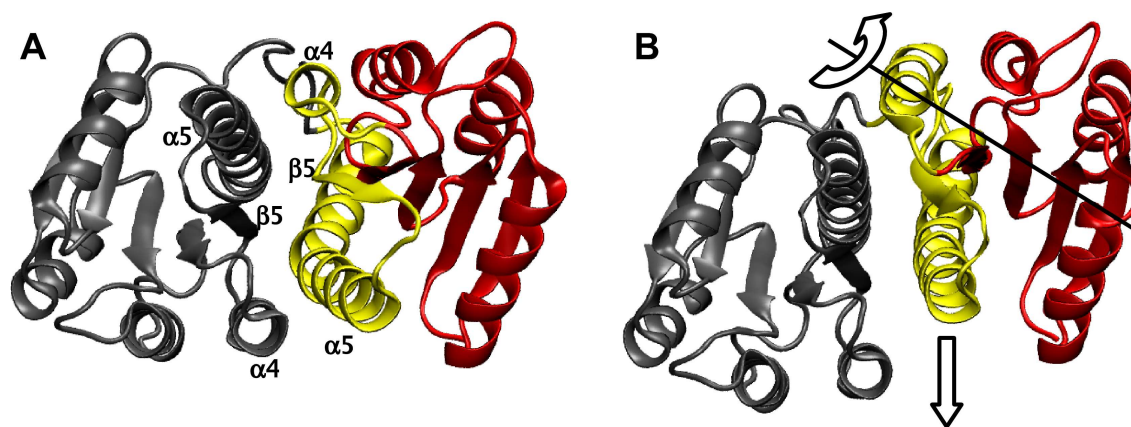
#### $\alpha 4$ - $\beta 5$ - $\alpha 5$ modulated interface area

Comparison of the experimental structures revealed a substantial reorientation of the two receiver domains upon phospho-activation which results in a two-fold symmetric domain arrangement (see Figure 8). This rearrangement within the monomer is also reflected by different sizes of the contact area between domains D1 and D2. Calculation of the buried solvent accessible surface area (SASA) shows an increase from 2,700 Å<sup>2</sup> to 3,500 Å<sup>2</sup> indicating a more optimal domain interface in the active state. Interestingly, in the *adimer* simulation, a sharp decrease of 300 Å<sup>2</sup> in the interaction area is observed at 2.2 and 2.6 ns for monomer A and B, respectively, which is approximately 0.5 ns after the  $\beta 4$ - $\alpha 4$  loop relaxation in monomer B. This event needs to be investigated further to draw any conclusions on whether loop repositioning is related to the interface change.

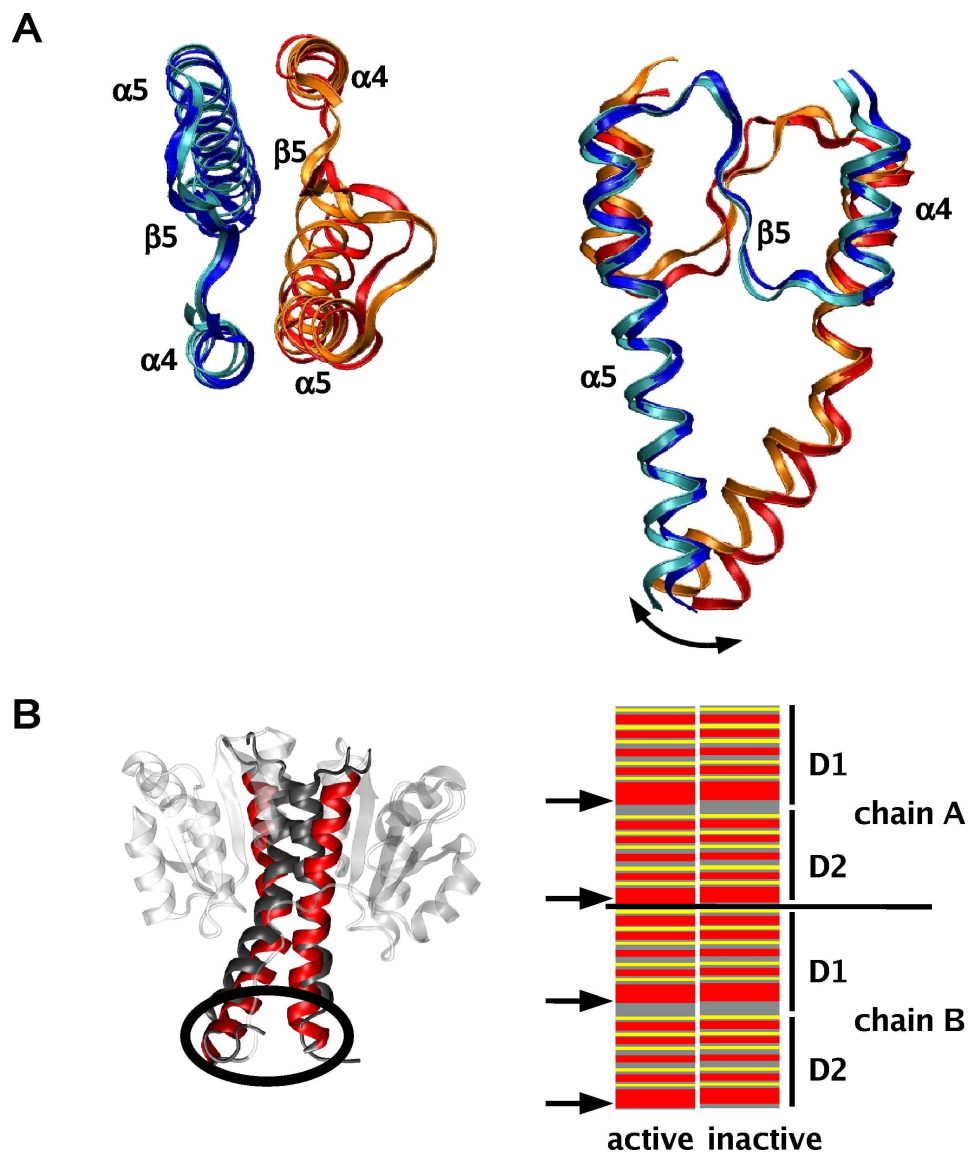
**Table 6: Characteristics of phospho-activation in CheY-like RRs**

	active	inactive	during MD simulation
<b><math>\beta</math>4-<math>\alpha</math>4 loop conformation</b>			
$C_{\alpha}$ 83-86	104° / 100°	-91° / -91°	transitions observed
$C_{\alpha}$ 232-235	-136° / -146°	-142° / -146°	unchanged
<b>F102 sidechain orientation</b>			
$\chi_1$ F102	-177° / -179°	-58° / -57°	unchanged, except monomer A ( <i>active</i> ): $t \rightarrow g^-$ at 2.5 ns
$\chi_1$ I251	-80° / -37°	47° / -28°	inactive: unchanged active: $t, g^-$ and $g^+$ sampled
<b>T83 sidechain orientation</b>			
$\chi_1$ T83	173° / 172°	59° / 60°	unchanged (active: $t$ , inactive: $g^-$ )
$\chi_1$ V232	-63° / -59°	-59° / -62°	unchanged, except monomer B (A_Y26A): $t \rightarrow g^- \rightarrow t$ at 1.2 and 1.3 ns

All values are indicated for monomer A and B, respectively. Corresponding residues in domain D2 from alignment in ref. 136.

**Figure 8: Reorientation of domains D1 and D2**

Cartoon representation of the D1/D2 interface in the (A) active and (B) inactive X-ray structure. The  $\alpha$ 4- $\beta$ 5- $\alpha$ 5 face of domain D1 (red) is highlighted (yellow), while domain D2 is colored grey. Arrows and axis indicate the reorientation of domain D1 with respect to D2 upon activation. Note the symmetry in the active state.

Extended helix  $\alpha 5$  has leverage effectFigure 9: Motion of the extended helix  $\alpha 5$ 

**A)** Representative structures for the active (red, blue) and inactive (orange, cyan) state. Rearrangements at the  $\alpha 4$ - $\beta 5$ - $\alpha 5$  face lead to large shifts at the C-terminal end of the helix  $\alpha 5$ . **B)** Domains D1 and D2 in cartoon representation with helix  $\alpha 5$  indicated in the activated (red) and inactive (grey) state. Secondary structure annotation obtained with the module STRIDE<sup>174</sup> implemented in VMD<sup>175</sup> for the two X-ray structures.  $\alpha$ -helices colored in red,  $\beta$ -strands in yellow and unstructured parts in grey.

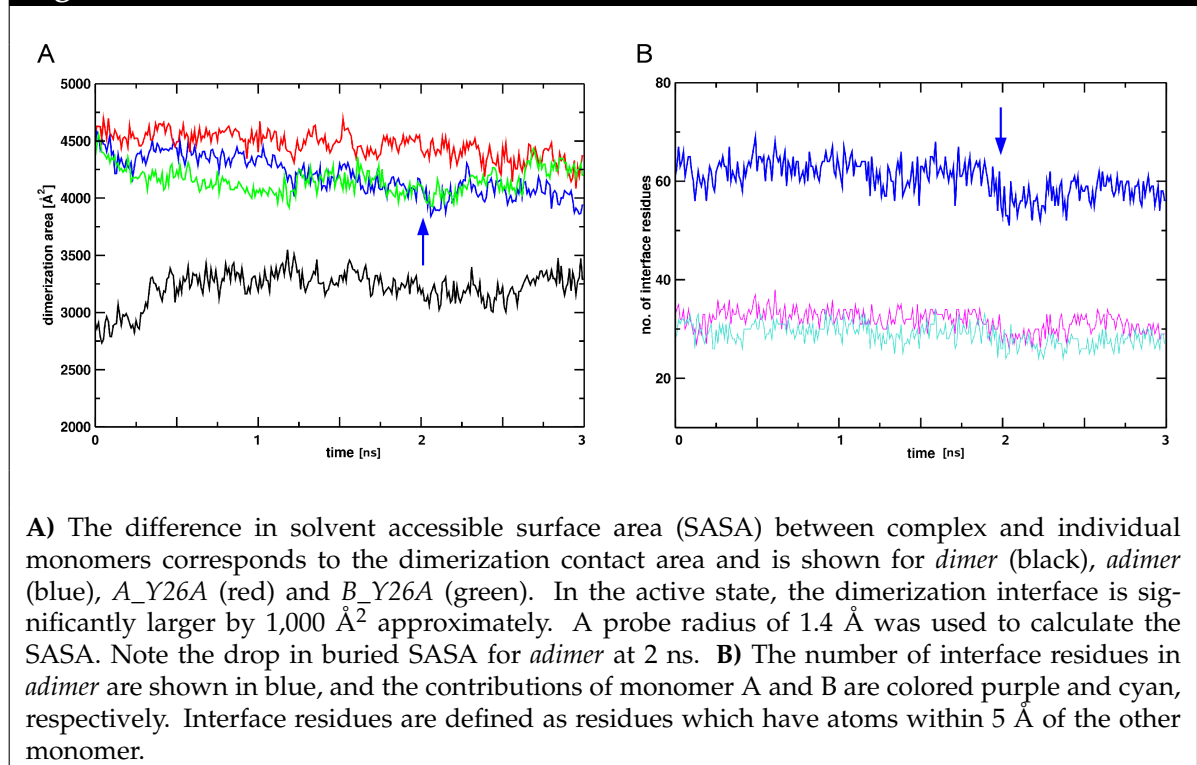
To study structural differences within one monomer that were induced by the active-to-inactive transition of the  $\beta 4$ - $\alpha 4$  loop, snapshots taken every 100 ps were super-

imposed on the backbone atoms of domain D2. Two distinct conformations are observed, one for the time before and one after loop repositioning (see Figure 9 A). Rearrangements in the  $\alpha 4$ - $\beta 5$ - $\alpha 5$  face were observed, e.g. displacement of the  $\beta 5$  strand. Yet, largest shifts are found for the C-terminal end of the extended helix  $\alpha 5$ . The  $\alpha 4$ - $\beta 5$ - $\alpha 5$  surface is well known as adaptive region, for example binding of CheY to FliM<sup>168</sup> or in homodimerization of the OmpR/PhoB response regulator subfamily<sup>176-179</sup> which is very similar to PleD. In the latter the two-fold symmetric interaction is promoted by phosphorylation and specific H-bonds and salt-bridges formed between highly conserved residues. Several of these residues are also found in PleD. However, common RR domains have a globular structure whereas domains D1 and D2 of PleD are "P-shaped". The main globular part of the domain is located at the N-terminal end of the unique extended helix  $\alpha 5$ . Based on the result that the  $\alpha 5$  C-terminal end is displaced most when the  $\beta 4$ - $\alpha 4$  loop is relaxed the following model can be suggested. A slight domain adjustment at the adaptive interface  $\alpha 4$ - $\beta 5$ - $\alpha 5$  is amplified towards the C-terminal end by the remarkable length of the helix  $\alpha 5$  in a leverage fashion.

Furthermore, in the active state the long helix  $\alpha 5$  is extended at the C-terminal end by 4-5 residues which correspond to a little more than one turn (see Figure 9 B). The secondary structure was predicted for snapshots taken every 10 ps using the STRIDE module<sup>174</sup> implemented in VMD<sup>175</sup>. Overall, the structure is preserved in all simulations. The least stable is the region around helix  $\alpha 3$  which is sometimes assigned as turn,  $\alpha$ - or  $\alpha_{3\_10}$ -helix.

### Dimerization interface

It is found from the X-ray structures that PleD in the  $\text{BeF}_3^-$  activated form has a more optimal dimerization interface reflected in an increase from 900  $\text{\AA}^2$  to 1,436  $\text{\AA}^2$  per monomer.<sup>52</sup> The contact area is calculated as the buried solvent accessible surface area (SASA) using a probe of the radius 1.4  $\text{\AA}$ . A difference of 1,000  $\text{\AA}^2$  is also calculated from the simulations (see Figure 10 A black vs. blue, red and green line). The Y26A mutation is thought to decrease the dimerization area due to the smaller volume of the residue sidechain. However, a similar area is observed for the mutant Y26A\_B and an even larger value is found for Y26A\_A. In the *adimer* simulation, a dip from 4,400 to 4,000  $\text{\AA}^2$  is detected around 2 ns. Interestingly, this event occurs shortly after the  $\beta 4$ - $\alpha 4$  loop relaxation. Calculation of the interface residues, i.e. residues

**Figure 10: Dimer contact area and interface residues**

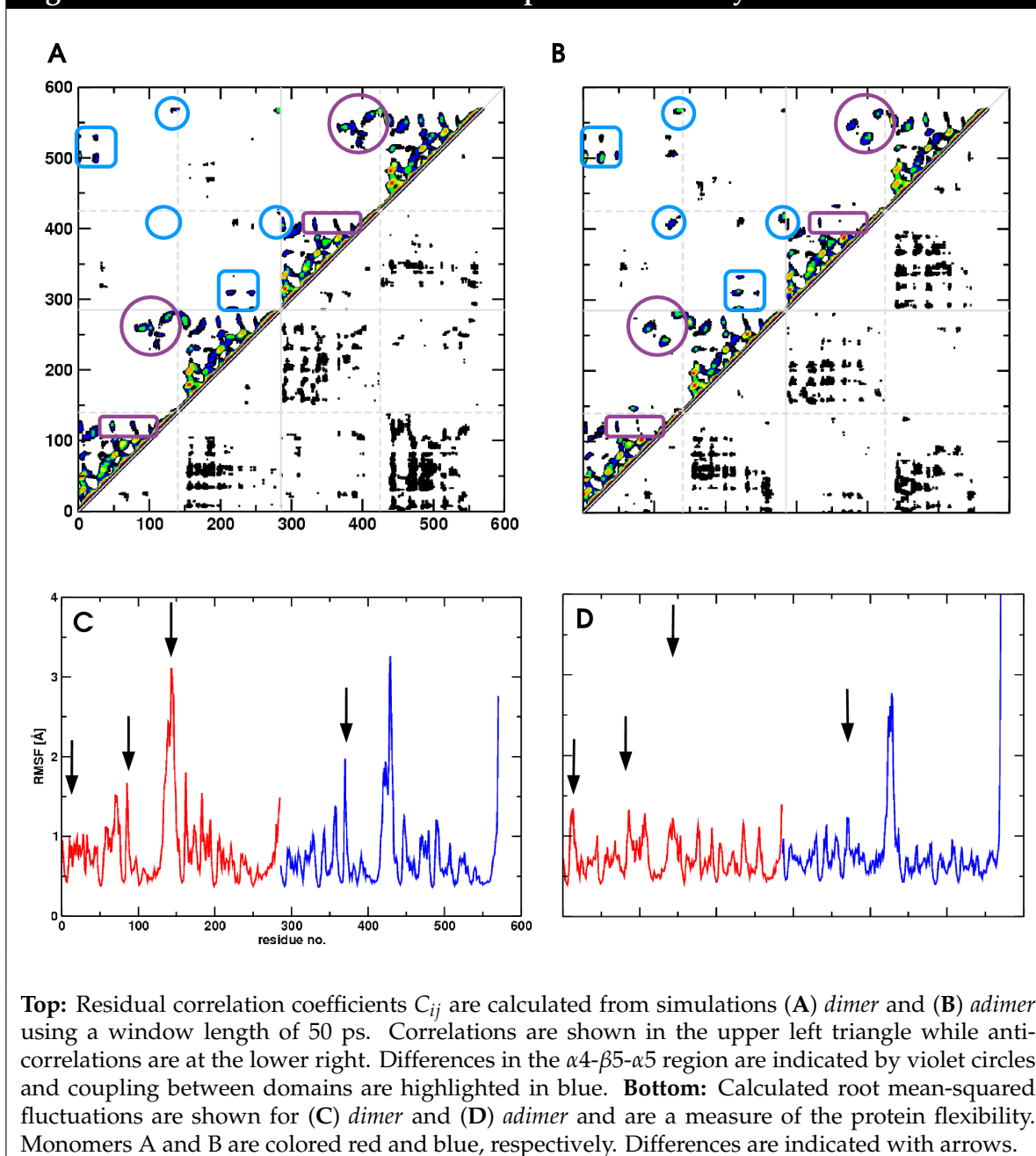
which have at least one atom within a distance of 5 Å of the other monomer, showed the loss of 5 interface residues in each monomer simultaneously (see Figure 10 B). The fact that the buried SASA recovers partly after the structural change argues for adaptation of the interacting surfaces similar to the induced fit phenomenon. Furthermore, in simulation *B\_Y26A* monomer B responds 0.5 ns after changes in monomer A which is indicated by the number of contact residues (data not shown).

## System dynamics

### Cross-correlated motions

Structure elements of the protein system which move in a coupled fashion are often depicted in dynamical cross-correlation maps (DCCMs). Residual correlation coefficients  $C_{ij}$  between residues  $i$  and  $j$  are calculated from the 3 ns trajectories and can adopt a value between -1 and 1 corresponding to anti-correlation and correlation. Persistent correlations are interesting because they can potentially transmit signals over long-range distances, in particular if correlations are found between different domains or subunits. In Figure 11 A and B the DCCM are shown for *dimer* and

Figure 11: Cross-correlated motion and protein flexibility



*adimer*. The DCCM for mutants *A\_Y26A* and *B\_Y26A* are similar to *adimer* and not shown here. Overall, the pattern within the individual domains shows little changes between the active and inactive conformation. In contrast, couplings between different domains and subunits are enhanced or newly formed in the active form. Most prominent differences concern interactions with helix  $\alpha 5$  (marked violet in Figure 11 A and B). In the active state the core  $\beta$ -sheet is less coupled to  $\alpha 5$  in domain D1 but

## 5 Results

---

not in D2. In addition changes in the D1/D2 interface  $\alpha 4$ - $\beta 5$ - $\alpha 5$  are observed, where  $\beta 5$  is no longer correlated to helix  $\alpha 5$  of the other domain when activated. However, correlations between helices  $\alpha 4$  and  $\alpha 5$  remain. Correlations between the monomers (marked blue in Figure 11 A and B) are mainly through interactions of the globular part of the protein and additional correlations between the  $\alpha 5$  helices in the active state reflecting the tight helix bundle structure. Thus, results from cross-correlated motion analysis also indicate the special role of the extended helix  $\alpha 5$  in that it interacts with the different domains and subunit, and differs in the modulated  $\alpha 4$ - $\beta 5$ - $\alpha 5$  area between the active and inactive state.

### Root mean-square fluctuations

The protein flexibility can be characterized by root mean-square fluctuations (RMSFs) calculated from the trajectory, where large RMSF values indicate highly mobile regions. In general, the RMSFs are smaller in the active than in the inactive state (see Figure 11 C and D), stating that the protein in the active conformation is overall less flexible. In detail, prominent differences are observed for the region linking domains D1 and D2 in monomer A (residues 130-150) being more rigid in *adimer*. This difference in amplitude between monomer A and B introduces some sort of asymmetry to the dimer system. Moreover, the  $\beta 4$ - $\alpha 4$  loop residues are more flexible in the inactive form (residues 83-86 and correspondingly at position 368-371). This finding may be of interest since the  $\beta 4$ - $\alpha 4$  loop is part of the functional activation process. In the mutant systems the previously described features resemble more the inactive state. In contrast, the flexibility of helix  $\alpha 1$  (residues 13-25) in domain D1 is enhanced in *adimer*, while in all other simulations this part is slightly more rigid. Regarding the mutant simulations this is somewhat surprising since the replacement of Tyrosine by the small Alanine is thought to decrease interactions and thus provide more flexibility.

Even though differences between simulations of the active and inactive system are found in both the cross-correlated motions and protein flexibility, 3 ns trajectories most probably are not sufficient to capture the entire dynamics intrinsic to the system and longer simulation times are needed.



## Conclusions & Outlook

Molecular dynamics simulations have the advantage of atomistic resolution, inherent high time resolution and the possibility to study species not characterized thus far by experiments. The present work shows that it is possible to observe and analyze the active-to-inactive transition in PleD within 3 ns simulation time. The  $\beta 4$ - $\alpha 4$  loop relaxation is found to be faster in monomeric PleD arguing for a stabilization effect from the associated monomer. In addition, the loop repositioning leads to structural rearrangements in the D1/D2 interface, particularly involving the  $\alpha 4$ - $\beta 5$ - $\alpha 5$  region. Based on structure comparison the following model was proposed: Small angle reorientations at the N-terminal end of the extended helix  $\alpha 5$  are amplified towards the C-terminal end, having a leverage effect. Helix  $\alpha 5$  has strong correlations to other domains and subunits corroborating its importance to protein function. However, no conclusions on how the contact residue Y26 mediates dimerization can be drawn and further simulations are necessary. For example, calculation of residual interaction energies can clarify if residue Y26 is a "hot-spot" residue, a residue contributing significantly to the overall interaction energy. Furthermore, biased simulations can be envisaged. Steered MD where the monomers are pulled apart can give information about which contacts are broken first and may also reveal interaction networks and the functional role of residue Y26.

Another interesting aspect to study is the effect from the associated monomer, for example whether it is able to activate its partner. MD simulations are especially suited to analyze mixed dimers, i.e. an activated monomer interacting with an inactive one or wildtype PleD with mutants. Genetics experiments discovered a constitutively active PleD mutant<sup>147,157,158</sup>, PleD\*, which is locked in the active state and is predominantly present as dimers in solution.

## Acknowledgments

Generous allocation of computing time at the CSCS in Manno, Switzerland is acknowledged.



## 6 Discussion

Extensive research on the small signaling molecule cyclic diguanosine monophosphate (c-d-GMP) has highlighted its central role in bacterial metabolism (see reviews<sup>141,142</sup>). Mutation and/or overexpression experiments combined with phenotypic evaluation linked c-di-GMP turnover with proteins containing GGDEF and EAL domains and demonstrated the implication of the signaling molecule in cell motility, virulence and biofilm formation. Experimental results from co-workers showed that the c-di-GMP concentration is tightly regulated *via* feedback inhibition, a property intrinsic to the GGDEF domain<sup>148</sup>, and that the response regulator PleD is additionally controlled by phosphorylation-mediated dimerization<sup>52,158</sup>. Based on PleD X-ray structures a "domain immobilization" mechanism was proposed, where contact between the active sites and subsequent c-di-GMP production is prevented.<sup>51,52</sup> Results from the present work contribute to the mechanistic understanding of DGC activity regulation, in particular the system's dynamics and effects from ligand-binding and phospho-activation.

### Communication between allosteric and active site

Comparison of energy optimized structures of monomeric PleD with and without c-di-GMP present in the allosteric site have indicated a balancelike movement of the  $\beta$ -strand  $\beta_2$ , that physically connects the active with the allosteric site.<sup>148</sup> Thus, ligand binding repositioning the I-site loop may potentially lead to misalignment of the enzymatic residues and prevent catalysis. This structure based mechanism works not only for PleD, but can be applied to the GGDEF domain in general. In addition, a substantial drop in flexibility is observed for I- and A-site residues simultaneously, arguing for dynamical coupling.<sup>81</sup> This finding emphasizes the dynamic contribution to allostery. In order to elucidate dynamic signaling pathways, correlated motions were calculated from trajectories. Results suggest two possible communication cascades running either along nearby secondary structure elements, including the previously mentioned strand  $\beta_2$ , or across the disordered regions that link the neighboring

domains. It was also found that the presence or absence of the ligand alters the correlation between domains such that the interaction between D1 and D2 is enhanced in free PleD. This long-range effect is of potential interest since it is known that domains D1 and D2 rearrange upon activation.

### **Flexibility of the DGC domain vs. the stem domains**

Conformation and energetics of unliganded PleD were probed using high-temperature dynamics simulations and biased domain repositioning.<sup>81</sup> An "open" form of PleD, being considerably lower in energy than the starting X-ray structure where bound c-di-GMP was removed, is in accord with the recent X-ray structure of activated PleD<sup>52</sup>. The fact that both conformations, and possibly still more conformations are energetically stable corroborates the model of the DGC domain being flexibly attached to the D1/D2 stem. This aspect is the basis for the suggested "domain immobilization" model where ligand binding fixes one non-productive domain orientation. However, nothing is known about the actual energy barriers for the domain to adopt different conformations.

### **Activation mechanism in view of the dimerization process**

Dynamics simulations were carried out using D1/D2 dimers in the inactive and  $\text{BeF}_3^-$ -activated conformation. Spontaneous relaxation of the  $\beta 4$ - $\alpha 4$  loop was observed within the 3 ns simulation time. This active-to-inactive transition is found to be correlated with rearrangements in the D1/D2 interface and leads to a slight decrease of the dimerization contact area. Specifically, adjustments in the  $\alpha 4$ - $\beta 5$ - $\alpha 5$  modulating region are amplified towards the C-terminal end of the extended helix  $\alpha 5$ . Moreover, this  $\alpha 5$  helix is implicated in interactions between different domains and subunits as indicated by cross-correlation maps.

### **Contribution of computational methods to understand protein function**

Molecular visualization of structural motion is an important element to build mechanistic models to explain protein function. Immense advances in computer technology now allow simulations approaching the  $\mu\text{s}$  regime and computational sciences are increasingly recognized to provide detailed insight into protein motion, residue interplay and energetics. In PleD, for example, dynamics information and potential communication routes were found using molecular simulations. Each of the methods reviewed has its own strengths and weaknesses with respect to sensitivity and

specificity emphasizing their complementarity. Thus, general consensus from many techniques, theoretical and experimental, increases the reliability of the proposed mechanism. Subdividing the complexity of supramolecular assemblies in biology into simple but quantitatively precise questions that can be solved by specific "experiments" should prove important. The focus of computational approaches lies in characterizing sequential events to understand how information is transferred.<sup>180</sup> Protein allostery surely is interdisciplinary, and brings together researchers with very distinct but complementary expertise in molecular biology, computational science, and physics. Easy-to-use web-based programs and standardized setups for molecular dynamics simulations could encourage experimentalists to interpret and complement their data with calculations, enabling a closer collaboration with theoreticians. In turn, simulation and programming experts might focus on force field improvements and the development of efficient algorithms.

### **Protein allostery: a constantly evolving concept**

The combined efforts of different disciplinary fields have provided new ideas and concepts on allosteric regulation over time. The two most widely accepted are the population-shift of co-existing states and the propagation of structural change along predefined pathways when triggered by effector binding. In addition, protein dynamics undoubtedly contribute to function and even allow allostery to take place without visible structural change.<sup>26,27,29</sup> All these views on protein allostery are not mutually exclusive and may be true to some extent. In PleD, suggested models involve structurally distinct states and dynamical structural pathways. It is possible that there exists a common mechanism that holds for all biomolecular systems, but which has not yet been found because the different features are differently pronounced, and may be even too small to detect with current methods. In any case, allosteric regulation is vital for a cell to respond to environmental signals, and detailed insight may be used for practical application in drug and protein design.<sup>181</sup>



## 7 Future Directions

### Characterization of binding mode & drug design

The design of potent DGC inhibitors is of substantial medical interest to combat infections caused by bacterial biofilms. For drug molecules to be successful, high affinity which allows medication in the nM to  $\mu$ M range and receptor specificity are most important. Computational methods to estimate binding free energies<sup>110–112</sup> can be used in the drug discovery process to discriminate binding ligands from non-binding ligands, and even rank the ligands according to their binding affinity. Though, it should be noted that the accuracy of prediction is important to be of pharmacological use. In contrast with experimental affinity measurements, computational studies have the advantage that they combine binding free energies with structural features. The individual contributions from chemical groups and the different interaction types are easily calculated. Structural properties such as sidechain conformations, salt-bridges, H-bonds and the size of the binding pocket can be combined into fingerprints and the subsequent use of clustering techniques may detect features that distinguish binders from non-binders. Such insight is valuable to rational drug design.

A suitable system to study c-di-GMP binding is the single domain protein DgcA of *C. crescentus* (CC3285) for which inhibition constants,  $K_i$ , have been measured for several I-site mutants.<sup>148</sup> In contrast to common ligand screening, here the ligand c-di-GMP stays the same but the protein binding site differs. Once the determinants of binding are known, computational and experimental ligand modification to increase the affinity to the allosteric site can be envisaged. So far, c-di-GMP binding sites contain a significant number of Arginine residues, in both PleD and the first identified c-di-GMP receptor domain, the PilZ domain<sup>153–155</sup>. Detection of important protein-ligand interactions in the DgcA system is relevant due to the fact that the conserved motif RxxD (residues 359 to 362 in PleD) required for inhibition is found in more than 60% of DGCs.

### **Force field extension to study protein phosphorylation**

Modulation of protein activity is often achieved by chemical modifications, whereof phosphotransfers between kinases and response regulators are most common forming "two-component" signaling pathways.<sup>162</sup> However, common force fields include parameters for phospho-groups for RNA and DNA molecules but mainly lack suitable phospho-aminoacid parameters. In the present study the problem of missing parameters is circumvented by removing the phospho-group (or  $\text{BeF}_3^-$  moiety) from the active conformation structure and analyzing the relaxation process from the active to the inactive form. However, for future studies it is desirable to have force field parameters which will allow the study of energetics. A first attempt to find parameters that reproduce geometries and energies from *ab initio* calculations has been made.<sup>182</sup> In addition, treating phosphorylation as a perturbation to the energy landscape proved useful to predict local structural changes, and helps to understand energetic components driving the structural rearrangement.<sup>183</sup>



# Bibliography

1. A. V. Hill. The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *J. Physiol. (Lond.)*, 40:iv–vii, 1910.
2. A. F. Cullis, H. Muirhead, M. F. Perutz, M. G. Rossmann, and A. C. T. North. The structure of haemoglobin. IX. a three-dimensional Fourier synthesis at 5.5 Å resolution: Description of the structure. *Proc. R. Soc. Lond. A*, 265:161–187, 1962.
3. M. F. Perutz. Stereochemistry of cooperative effects in haemoglobin. *Nature*, 228:726–734, 1970.
4. D. E. Koshland, Jr., G. Némethy, and D. Filmer. Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry*, 5:365–385, 1966.
5. J. Monod, J. Wyman, and J.-P. Changeux. On the nature of allosteric transitions: A plausible model. *J. Mol. Biol.*, 12:88–118, 1965.
6. B. F. Volkman, D. Lipson, D. E. Wemmer, and D. Kern. Two-state allosteric behavior in a single-domain signaling protein. *Science*, 291:2429–2433, 2001.
7. J. Gsponer, J. Christodoulou, A. Cavalli, J. M. Bui, B. Richter, C. M. Dobson, and M. Vendruscolo. A coupled equilibrium shift mechanism in calmodulin-mediated signal transduction. *Structure*, 16:736–746, 2008.
8. J. N. Onuchic, Z. Luthey-Schulten, and P. G. Wolynes. Theory of protein folding: The energy landscape perspective. *Annu. Rev. Phys. Chem.*, 48:545–600, 1997.
9. J. N. Onuchic and P. G. Wolynes. Theory of protein folding. *Curr. Op. Struct. Biol.*, 14:70–75, 2004.
10. D. Kern and E. R. P. Zuiderweg. The role of dynamics in allosteric regulation. *Curr. Op. Struct. Biol.*, 13:748–757, 2003.

11. K. Henzler-Wildman and D. Kern. Dynamic personalities of proteins. *Nature*, 450:964–972, 2007.
12. A. Mittermaier and L. E. Kay. New tools provide new insights in NMR studies of protein dynamics. *Science*, 312:224–228, 2006.
13. M. Akke. NMR methods for characterizing microsecond to millisecond dynamics in recognition and catalysis. *Curr. Op. Struct. Biol.*, 12:642–647, 2002.
14. A. G. Palmer III, C. D. Kroenke, and J. P. Loria. Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol.*, 339:204–238, 2001.
15. M. Diez, B. Zimmermann, M. Börsch, M. König, E. Schweinberger, S. Steigmiller, R. Reuter, S. Felekyan, V. Kudryavtsev, C. A. M. Seidel, and P. Gräber. Proton-powered subunit rotation in single membrane-bound  $F_0F_1$ -ATP synthase. *Nat. Struct. Mol. Biol.*, 11:135–141, 2004.
16. F. Schotte, M. Lim, T. A. Jackson, A. V. Smirnov, J. Soman, J. S. Olson, G. N. Phillips Jr., M. Wulff, and P. A. Anfinrud. Watching a protein as it functions with 150-ps time-resolved X-ray crystallography. *Science*, 300:1944–1947, 2004.
17. V. Šrajer, Z. Ren, T.-Y. Teng, M. Schmidt, T. Ursby, D. Bourgeois, C. Pradervand, W. Schildkamp, M. Wulff, and K. Moffat. Protein conformational relaxation and ligand migration in myoglobin: A nanosecond to millisecond molecular movie from time-resolved Laue X-ray diffraction. *Biochemistry*, 40:13802–13815, 2001.
18. J. E. Knapp, R. Pahl, V. Šrajer, and W. E. Royer Jr. Allosteric action in real time: Time-resolved crystallographic studies of a cooperative dimeric hemoglobin. *Proc. Natl. Acad. Sci. USA*, 103:7649–7654, 2006.
19. A. Liwo, C. Czaplewski, S. Oldziej, and H. A. Scheraga. Computational techniques for efficient conformational sampling of proteins. *Curr. Op. Struct. Biol.*, 18:134–139, 2008.
20. H. Valadié, J. J. Lacapère, Y.-H. Sanejouand, and C. Etchebest. Dynamical properties of the MscL of *Escherichia coli*: A normal mode analysis. *J. Mol. Biol.*, 332:657–674, 2003.

21. Q. Cui, G. Li, J. Ma, and M. Karplus. A normal mode analysis of structural plasticity in the biomolecular motor F<sub>1</sub>-ATPase. *J. Mol. Biol.*, 340:345–372, 2004.
22. G. Li and Q. Cui. Analysis of functional motions in brownian molecular machines with an efficient block normal mode approach: Myosin-II and Ca<sup>2+</sup>-ATPase. *Biophys. J.*, 86:743–763, 2004.
23. M Lu and J. Ma. The role of shape in determining molecular motions. *Biophys. J.*, 89:2395–2401, 2005.
24. J. Ma. Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. *Structure*, 13:373–380, 2005.
25. K. Lindorff-Larsen, R. B. Best, M. A. DePristo, C. M. Dobson, and M. Vendruscolo. Simultaneous determination of protein structure and dynamics. *Nature*, 433:128–132, 2005.
26. C.-J. Tsai, A. del Sol, and R. Nussinov. Allostery: Absence of a change in shape does not imply that allostery is not at play. *J. Mol. Biol.*, 378:1–11, 2008.
27. N. Popovych, S. Sun, R. H. Ebright, and C. G. Kalodimos. Dynamically driven protein allostery. *Nat. Struct. Mol. Biol.*, 13:831–838, 2006.
28. A. Cooper and D. T. F. Dryden. Allostery without conformational change. *Eur. Biophys. J.*, 11:103–109, 1984.
29. K. Gunasekaran, B. Ma, and R. Nussinov. Is allostery an intrinsic property of *all* dynamic proteins? *Proteins: Struct. Funct. Bioinf.*, 57:433–443, 2004.
30. K. A. Henzler-Wildman, V. Thai, M. Lei, M. Ott, M Wolf-Watz, T. Fenn, E. Pozharski, M. A. Wilson, G. A. Petsko, M. Karplus, C. G. Hübner, and D. Kern. Intrinsic motions along an enzymatic reaction trajectory. *Nature*, 450:838–844, 2007.
31. F. Tama and Y.-H. Sanejouand. Conformational change of proteins arising from normal mode calculations. *Protein Engineering*, 14:1–6, 2001.
32. D. Tobi and I. Bahar. Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. *Proc. Natl. Acad. Sci. USA*, 102:18908–18913, 2005.

33. S. W. Lockless and R. Ranganathan. Evolutionarily conserved pathways of energetic connectivity in protein families. *Science*, 286:295–299, 1999.
34. G. M. Süel, S. W. Lockless, M. A. Wall, and R. Ranganathan. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nat. Struct. Biol.*, 10:59–68, 2003.
35. N. Ota and D. A. Agard. Intramolecular signaling pathways revealed by modeling anisotropic thermal diffusion. *J. Mol. Biol.*, 351:345–354, 2005.
36. K. Sharp and J. J. Skinner. Pump-probe molecular dynamics as a tool for studying protein motion and long range coupling. *Proteins: Struct. Funct. Bioinf.*, 65:347–361, 2006.
37. M. S. Formanek, L. Ma, and Q. Cui. Reconciling the "old" and "new" views of protein allostery: A molecular simulation study of chemotaxis Y protein (CheY). *Proteins: Struct. Funct. Bioinf.*, 63:846–867, 2006.
38. B. Isralewitz, M. Gao, and K. Schulten. Steered molecular dynamics and mechanical functions of proteins. *Curr. Op. Struct. Biol.*, 11:224–230, 2001.
39. J. Ma, P. B. Sigler, Z. Xu, and M. Karplus. A dynamic model for the allosteric mechanism of GroEL. *J. Mol. Biol.*, 302:303–313, 2000.
40. B. Isralewitz, S. Izrailev, and K. Schulten. Binding pathway of retinal to bacteriorhodopsin: A prediction by molecular dynamics simulations. *Biophys. J.*, 73:2972–2979, 1997.
41. N. M. Goodey and S. J. Benkovic. Allosteric regulation and catalysis emerge via a common route. *Nat. Chem. Biol.*, 4:474–482, 2008.
42. Q. Cui and M. Karplus. Allostery and cooperativity revisited. *Prot. Sci.*, 17:1–13, 2008.
43. J. Kuriyan and D. Eisenberg. The origin of protein interactions and allostery in colocalization. *Nature*, 450:983–990, 2007.
44. A. Fersht. The basic equations of enzyme kinetics. In *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*, A. Fersht, Ed., W.H. Freeman, 1998

45. H. Lineweaver and D. Burk. The determination of enzyme dissociation constants. *J. Am. Chem. Soc.*, 56:658–666, 1934.
46. B. H. J. Hofstee. Non-inverted versus inverted plots in enzyme kinetics. *Nature*, 184:1296–1298, 1959.
47. E. M. Phizicky and S. Fields. Protein-protein interactions: Methods for detection and analysis. *Microbiol. Rev.*, 59:94–123, 1995.
48. S. Fields and O.-K. Song. A novel genetic system to detect protein-protein interactions. *Nature*, 340:245–246, 1989.
49. A. Sinz. Chemical cross-linking and mass spectrometry for mapping three-dimensional structures of proteins and protein complexes. *J. Mass Spectrom.*, 38:1225–1237, 2003.
50. S.-Y. Lee, H. S. Cho, J. G. Pelton, D. Yan, E. A. Berry, and D. E. Wemmer. Crystal structure of activated CheY. *J. Biol. Chem.*, 276:16425–16431, 2001.
51. C. Chan, R. Paul, D. Samoray, N. C. Amiot, B. Giese, U. Jenal, and T. Schirmer. Structural basis of activity and allosteric control of diguanylate cyclase. *Proc. Natl. Acad. Sci. USA*, 101:17084–17089, 2004.
52. P. Wassmann, C. Chan, R. Paul, A. Beck, H. Heerklotz, U. Jenal, and T. Schirmer. Structure of BeF<sub>3</sub><sup>-</sup>-modified response regulator PleD: Implications for diguanylate cyclase activation, catalysis, and feedback inhibition. *Structure*, 15:915–927, 2007.
53. K. Moffat. Time-resolved biochemical crystallography: A mechanistic perspective. *Chem. Rev.*, 101:1569–1581, 2001.
54. R. C. Wilmoth, I. J. Clifton, and R. Neutze. Recent successes in time-resolved protein crystallography. *Nat. Prod. Rep.*, 17:527–533, 2000.
55. H. R. Saibil. Macromolecular structure determination by cryo-electron microscopy. *Acta Cryst.*, D56:1215–1222, 2000.
56. J. Elands and W. Hax. CryoEM as a complement to current techniques in protein structural analysis. *Current Drug Discovery*, October:15–20, 2004.

57. A. Sali, R. Glaeser, T. Earnest, and W. Baumeister. From words to literature in structural proteomics. *Nature*, 422:216–225, 2003.
58. A. G. Palmer III. NMR probes of molecular dynamics: Overview and comparison with other techniques. *Annu. Rev. Biophys. Biomol. Struct.*, 30:129–155, 2001.
59. R. Das, M. Abu-Abed, and G. Melacini. Mapping allostery through equilibrium perturbation NMR spectroscopy. *J. Am. Chem. Soc.*, 128:8406–8407, 2006.
60. W. Jahnke. Spin labels as a tool to identify and characterize protein-ligand interactions by NMR spectroscopy. *Chem. Bio. Chem.*, 3:167–173, 2002.
61. P. Dosset, J.-C. Hus, D. Marion, and M. Blackledge. A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings. *J. Biomol. NMR*, 20:223–231, 2001.
62. G. Bouvignies, P. Bernadó, S. Meier, K. Cho, S. Grzesiek, R. Brüschweiler, and M. Blackledge. Identification of slow correlated motions in proteins using residual dipolar and hydrogen-bond scalar couplings. *Proc. Natl. Acad. Sci. USA*, 102:13885–13890, 2005.
63. F. Cordier and S. Grzesiek. Direct observation of hydrogen bonds in proteins by interresidue  $^3\text{h}J_{\text{NC}'}$  scalar couplings. *J. Am. Chem. Soc.*, 121:1601–1602, 1999.
64. M. Barfield. Structural dependencies of interresidue scalar coupling  $^3\text{h}J_{\text{NC}'}$  and donor  $^1\text{H}$  chemical shifts in the hydrogen bonding regions of proteins. *J. Am. Chem. Soc.*, 124:4158–4168, 2002.
65. S. A. Showalter and R. Brüschweiler. Validation of molecular dynamics simulations of biomolecules using NMR spin relaxation as benchmarks: Application to the AMBER99SB force field. *J. Chem. Theory Comput.*, 3:961–975, 2007.
66. F. F.-F. Schmid and M. Meuwly. Direct comparison of experimental and calculated NMR scalar coupling constants for force field validation and adaptation. *J. Chem. Theory Comput.*, 4:1949–1958, 2008.
67. A. A. Deniz, S. Mukhopadhyay, and E. A. Lemke. Single-molecule biophysics: at the interface of biology, physics and chemistry. *J. R. Soc. Interface*, 5:15–45, 2008.
68. T. Förster. Zwischenmolekulare Energiewanderung und Fluoreszenz. *Ann. Physik*, 6:55–75, 1948.

69. C. Joo, H. Balci, Y. Ishitsuka, C. Buranachai, and T. Ha. Advances in single-molecule fluorescence methods for molecular biology. *Annu. Rev. Biochem.*, 77:51–76, 2008.
70. P. R. Selvin. The renaissance of fluorescence resonance energy transfer. *Nat. Struct. Biol.*, 7:730–734, 2000.
71. M. Karplus and J. Kuriyan. Molecular dynamics and protein function. *Proc. Natl. Acad. Sci. USA*, 102:6679–6685, 2005.
72. A. W. van Wynsberghe, L. Ma, X. Chen, and Q. Cui. Functional motions in biomolecules: Insights from computational studies at multiple scales. In *Computational Structural Biology Methods and Applications*, T. Schwede, M. Peitsch, Eds., World Scientific Publishing Co. Pte. Ltd., 2008.
73. S. A. Adcock and J. A. McCammon. Molecular dynamics: Survey of methods for simulating the activity of proteins. *Chem. Rev.*, 106:1589–1615, 2006.
74. A. D. MacKerell, Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. E. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, 102:3586, 1998.
75. T. C. Bruice and K. Kahn. Computational enzymology. *Curr. Opin. Chem. Biol.*, 4:540–544, 2000.
76. G. Monard, X. Prat-Resina, A. González-Lafont, and J. M. Lluch. Determination of enzymatic reaction pathways using QM/MM methods. *Int. J. Quantum Chem.*, 93:229–244, 2003.
77. T. A. Halgren and W. Damm. Polarizable force fields. *Curr. Op. Struct. Biol.*, 11:236–242, 2001.
78. S. Lammers, S. Lutz, and M. Meuwly. Reactive force fields for proton transfer dynamics. *J. Comput. Chem.*, 29:1048–1063, 2008.

79. J. Danielsson and M. Meuwly. Atomistic simulation of adiabatic reactive processes based on multi-state potential energy surfaces. *J. Chem. Theory Comput.*, 4:1083–1093, 2008.
80. P. L. Freddolino, A. S. Arkhipov, S. B. Larson, A. McPherson, and K. Schulten. Molecular dynamics simulations of the complete satellite tobacco mosaic virus. *Structure*, 14:437–449, 2005.
81. F. F.-F. Schmid and M. Meuwly. All-atom simulations of structures and energetics of c-di-GMP-bound and free PleD. *J. Mol. Biol.*, 374:1270–1285, 2007.
82. M. Król, I. Roterman, B. Piekarska, L. Konieczny, J. Rybarska, B. Stopa, and P. Spólnik. Analysis of correlated domain motions in IgG light chain reveals possible mechanisms of immunological signal transduction. *Proteins: Struct. Funct. Bioinf.*, 59:545–554, 2005.
83. A. W. van Wynsberghe and Q. Cui. Interpreting correlated motions using normal mode analysis. *Structure*, 14:1647–1653, 2006.
84. P. De Los Rios, F. Cecconi, A. Pretre, G. Dietler, O. Michielin, F. Piazza, and B. Juanico. Functional dynamics of PDZ binding domains: A normal-mode analysis. *Biophys. J.*, 89:14–21, 2005.
85. A. Taly, M. Delarue, T. Grutter, M. Nilges, N. Le Novère, P.-J. Corringer, and J.-P. Changeux. Normal mode analysis suggests a quaternary twist model for the nicotinic receptor gating mechanism. *Biophys. J.*, 88:3954–3965, 2005.
86. S. Hayward. Normal mode analysis of biological molecules. In *Computational Biochemistry and Biophysics*, O. M. Becker, A. D. MacKerell Jr., B. Roux, and M. Watanabe, Eds., Routledge, USA, 2001.
87. Authors. Title. *Biopolymers*, 33:599, 1993.
88. D. Perahia and L. Mouawad. Computation of low-frequency normal modes in macromolecules: Improvements to the method of diagonalization in a mixed basis and application to hemoglobin. *Computers Chem.*, 19:241–246, 1995.
89. G. Li and Q. Cui. A coarse-grained normal mode approach for macromolecules: An efficient implementation and application to Ca<sup>2+</sup>-ATPase. *Biophys. J.*, 83:2457–2474, 2002.



90. O. Kurkcuoglu, R. L. Jernigan, and P. Doruker. Mixed levels of coarse-graining of large proteins using elastic network model succeeds in extracting the slowest motions. *Polymer*, 45:649–657, 2004.
91. K. Suhre and Y.-H. Sanejouand. ElNémo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic Acids Res.*, 32:610–614, 2004.
92. W. Zheng and B. R. Brooks. Modeling protein conformational changes by iterative fitting of distance constraints using reoriented normal modes. *Biophys. J.*, 90:4327–4336, 2006.
93. S. M. Hollup, G. Salensminde, and N. Reuter. WEBnm: a web application for normal mode analyses of proteins. *BMC Bioinformatics*, 6:52–59, 2005.
94. W. Zheng, B. R. Brooks, and G. Hummer. Protein conformational transitions explored by mixed elastic network models. *Proteins: Struct. Funct. Bioinf.*, 69:43–57, 2007.
95. S. Izrailev, S. Stepaniants, M. Balsera, Y. Oono, and K. Schulten. Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophys. J.*, 72:1568–1581, 1997.
96. C. Jarzynski. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.*, 78:2690–2693, 1997.
97. S. Park, F. Khalili-Araghi, E. Tajkhorshid, and K. Schulten. Free energy calculation from steered molecular dynamics simulations using Jarzynski’s equality. *J. Chem. Phys.*, 119:3559–3566, 2003.
98. N. Echols, D. Milburn, and M. Gerstein. MolMovDB: analysis and visualization of conformational change and structural flexibility. *Nucleic Acids Res.*, 31:478–482, 2003.
99. M. Gerstein and W. Krebs. A database of macromolecular motions. *Nucleic Acids Res.*, 26:4280–4290, 1998.
100. G. Qi, R. Lee, and S. Hayward. A comprehensive and non-redundant database of protein domain movements. *Bioinformatics*, 21:2832–2838, 2005.

101. Y. Xu, D. Xu, and H. N. Gabow. Protein domain decomposition using a graph-theoretic approach. *Bioinformatics*, 16:1091–1104, 2000.
102. S. Hayward and H. J. C. Berendsen. Systematic analysis of domain motions in proteins from conformational change: New results on citrate synthase and T4 lysozyme. *Proteins: Struct. Funct. Bioinf.*, 30:144–154, 1998.
103. S. Hayward and R. A. Lee. Improvements in the analysis of domain motions in proteins from conformational change: DynDom version 1.50. *J. Mol. Graphics Modell.*, 21:181–183, 2002.
104. K. Hinsen. Domain motions in proteins. *J. Molecular Liquids*, 84:53–63, 2000.
105. H. Gohlke, C. Kiel, and D. A. Case. Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RalGDS complexes. *J. Mol. Biol.*, 330:891–913, 2003.
106. V. Zoete, M. Meuwly, and M. Karplus. Study of the insulin dimerization from binding free energy calculations and per-residue free energy decomposition. *Proteins: Struct. Funct. Bioinf.*, 61:79–93, 2005.
107. I. S. Moreira, P. A. Fernandes, and M. J. Ramos. Hot spots – a review of the protein-protein interface determinant amino-acid residues. *Proteins: Struct. Funct. Bioinf.*, 68:803–812, 2007.
108. I. S. Moreira, P. A. Fernandes, and M. J. Ramos. Computational alanine scanning mutagenesis – an improved methodological approach. *J. Comput. Chem.*, 28:644–654, 2007.
109. I. S. Moreira, P. A. Fernandes, and M. J. Ramos. Unravelling hot spots: a comprehensive computational mutagenesis study. *Theor. Chem. Acc.*, 117:99–113, 2007.
110. H.-J. Woo and B. Roux. Calculation of absolute protein-ligand binding free energy from computer simulations. *Proc. Natl. Acad. Sci. USA*, 102:6825–6830, 2005.
111. J. Wang, Y. Deng, and B. Roux. Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials. *Biophys. J.*, 91:2798–2814, 2006.

112. C. Bartels, A. Widmer, and C. Ehrhardt. Absolute free energies of binding of peptide analogs to the HIV-1 protease from molecular dynamics simulations. *J. Comput. Chem.*, 26:1294–1305, 2005.
113. R. Abel, T. Young, R. Farid, B. J. Berne, and R. A. Friesner. Role of the active-site solvent in the thermodynamics of factor Xa ligand binding. *J. Am. Chem. Soc.*, 130:2817–2831, 2008.
114. M. A. Martí-Renom, A. C. Stuart, A. Fiser, R. Sánchez, M. Melo, and A. Šali. Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct*, 29:291–325, 2000.
115. D. Baker and A. Sali. Protein structure prediction and structural genomics. *Science*, 294:93–96, 2001.
116. T. Schwede, A. Sali, N. Eswar, and M. C. Peitsch. Protein structure modeling. In *Computational Structural Biology Methods and Applications*, T. Schwede, M. Peitsch, Eds., World Scientific Publishing Co. Pte. Ltd., 2008.
117. L. J. McGuffin. Protein fold recognition and threading. In *Computational Structural Biology Methods and Applications*, T. Schwede, M. Peitsch, Eds., World Scientific Publishing Co. Pte. Ltd., 2008.
118. S. E. Brenner. A tour of structural genomics. *Nat. Rev. Genet.*, 2:801–809, 2001.
119. M. R. Chance, A. R. Bresnick, S. K. Burley, J.-S. Jiang, C. D. Lima, A. Sali, S. C. Almo, J. B. Bonanno, J. A. Buglino, S. Boulton, H. Chen, N. Eswar, G. He, R. Huang, V. Ilyin, L. McMahan, U. Pieper, S. Ray, M. Vidal, and L. K. Wang. Structural genomics: A pipeline for providing structures for the biologist. *Prot. Sci.*, 11:723–738, 2002.
120. K. Arnold, L. Bordoli, J. Kopp, and T. Schwede. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*, 22:195–201, 2006.
121. J. D. Kahmann, H.-J. Sass, M. G. Allan, H. Seto, C. J. Thompson, and S. Grzesiek. Structural basis for antibiotic recognition by the TipA class of multidrug-resistance transcriptional regulators. *EMBO J.*, 22:1824–1834, 2003.

122. H. J. Dyson and P. E. Wright. Coupling of folding and binding for unstructured proteins. *Curr. Op. Struct. Biol.*, 12:54–60, 2002.
123. J. J. Ward, J. S. Sodhi, L. J. McGuffin, B. F. Buxton, and D. T. Jones. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.*, 337:635–645, 2004.
124. R. Linding, L. J. Jensen, F. Diella, P. Bork, T. J. Gibson, and R. B. Russell. Protein disorder prediction: Implications for structural proteomics. *Structure*, 11:1453–1459, 2003.
125. Z. R. Yang, R. Thomson, P. McNeil, and R. M. Esnouf. RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics*, 21:3369–3376, 2005.
126. R. Linding, R. B. Russell, V. Neduva, and T. J. Gibson. GlobPlot: exploring protein sequences for globularity and disorder. *Nucleic Acids Res.*, 31:3701–3708, 2003.
127. J. Gu, M. Gribskov, and P. E. Bourne. Wiggle - predicting functionally flexible regions from primary sequence. *PLoS Comp. Biol.*, 2:0769–0785, 2006.
128. M. Suyama and O. Ohara. DomCut: prediction of inter-domain linker regions in amino acid sequences. *Bioinformatics*, 19:673–674, 2003.
129. P. Ross, H. Weinhouse, Y. Aloni, D. Michaeli, P. Weinberger-Ohana, R. Mayer, S. Braun, E. de Vroom, G. A. van der Marel, J. H. van Boom, and M. Benziman. Regulation of cellulose synthesis in *Acetobacter xylinum* by cyclic diguanylic acid. *Nature*, 325:279–281, 1987.
130. U. Jenal. Cyclic di-guanosine-monophosphate comes of age: A novel secondary messenger involved in modulating cell surface structures in bacteria? *Curr. Op. Microbiol.*, 7:185–191, 2004.
131. D. A. Ryjenkov, M. Tarutina, O. V. Moskvin, and M. Gomelsky. Cyclic diguanylate is a ubiquitous signaling molecule in bacteria: Insights into biochemistry of the GGDEF protein domain. *J. Bacteriol.*, 187:1792–1798, 2005.
132. U. Römling and D. Amikam. Cyclic di-GMP as a second messenger. *Curr. Op. Microbiol.*, 9:218–228, 2006.

133. D. A. D'Argenio and S. I. Miller. Cyclic di-GMP as a bacterial second messenger. *Microbiology*, 150:2497–2502, 2004.
134. B. M Prüss, C. Besemann, A. Denton, and A. J. Wolfe. A complex transcription network controls the early stages of biofilm development by *Escherichia coli*. *J. Bacteriol.*, 188:3731–3739, 2006.
135. M. Y. Galperin. Bacterial signal transduction network in a genomic perspective. *Environ. Microbiol.*, 6:552–567, 2004.
136. G. B. Hecht and A. Newton. Identification of a novel response regulator required for the swarmer-to-stalked-cell transition in *Caulobacter crescentus*. *J. Bacteriol.*, 177:6223–6229, 1995.
137. P. Aldridge and U. Jenal. Cell cycle-dependent degradation of a flagellar motor component requires a novel-type response regulator. *Mol. Microbiol.*, 32:379–391, 1999.
138. B. García, C. Latasa, C. Solano, F. García-del Portillo C. Gamazo, and I. Lasa. Role of the GGDEF protein family in *Salmonella* cellulose biosynthesis and biofilm formation. *Mol. Microbiol.*, 54:264–277, 2004.
139. A. D. Tischler and A. Camilli. Cyclic diguanylate (c-di-GMP) regulates *Vibrio cholerae* biofilm formation. *Mol. Microbiol.*, 53:857–869, 2004.
140. P. A. Cotter and S. Stibitz. C-di-GMP-mediated regulation of virulence and biofilm formation. *Curr. Op. Microbiol.*, 10:17–23, 2007.
141. U. Jenal and J. Malone. Mechanisms of cyclic-di-GMP signaling in bacteria. *Annu. Rev. Genet.*, 40:385–407, 2006.
142. R. Tamayo, J. T. Pratt, and A. Camilli. Roles of cyclic diguanylate in the regulation of bacterial pathogenesis. *Annu. Rev. Microbiol.*, 61:131–148, 2007.
143. R. Simm, M. Morr, A. Kader, M. Nimtz, and U. Römling. GGDEF and EAL domains inversely regulate cyclic di-GMP levels and transition from sessility to motility. *Mol. Microbiol.*, 53:1123–1134, 2004.
144. R. Tal, H. C. Wong, R. Calhoon, D. Gelfand, A. L. Fear, G. Volman, R. Mayer, P. Ross, D. Amikam, H. Weinhouse, A. Cohen, S. Sapir, P. Ohana, and M. Benzi-man. Three *cdg* operons control cellular turnover of cyclic di-GMP in *Acetobacter*

- xylum*: Genetic organization and occurrence of conserved domains in isoenzymes. *J. Bacteriol.*, 180:4416–4425, 1998.
145. M. Y. Galperin, A. N. Nikolskaya, and E. V. Koonin. Novel domains of the prokaryotic two-component signal transduction systems. *FEMS Microbiol. Lett.*, 203:11–21, 2001.
146. J. Pei and N. V. Grishin. GGDEF domain is homologous to adenylyl cyclase. *Proteins: Struct. Funct. Bioinf.*, 42:210–216, 2001.
147. R. Paul, S. Weiser, N. C. Amiot, C. Chan, T. Schirmer, B. Giese, and U. Jenal. Cell cycle-dependent dynamic localization of a bacterial response regulator with a novel di-guanylate cyclase output domain. *Genes Dev.*, 18:715–727, 2004.
148. B. Christen, M. Christen, R. Paul, F. Schmid, M. Folcher, P. Jenoe, M. Meuwly, and U. Jenal. Allosteric control of cyclic di-GMP signaling. *J. Biol. Chem.*, 281:32015–32024, 2006.
149. N. Ausmees, R. Mayer, H. Weinhouse, G. Volman, D. Amikam, M. Benziman, and M. Lindberg. Genetic data indicate that proteins containing the GGDEF domain possess diguanylate cyclase activity. *FEMS Microbiol. Lett.*, 204:163–167, 2001.
150. M. Christen, B. Christen, M. Folcher, A. Schauerte, and U. Jenal. Identification and characterization of a cyclic di-GMP-specific phosphodiesterase and its allosteric control by GTP\*. *J. Biol. Chem.*, 280:30829–30837, 2005.
151. A. J. Schmidt, D. A. Ryjenkov, and M. Gomelsky. The ubiquitous protein domain EAL is a cyclic diguanylate-specific phosphodiesterase: Enzymatically active and inactive EAL domains. *J. Bacteriol.*, 187:4774–4781, 2005.
152. A. G. Bobrov, O. Kirillina, and R. D. Perry. The phosphodiesterase activity of the HmsP EAL domain is required for negative regulation of biofilm formation in *Yersinia pestis*. *FEMS Microbiol. Lett.*, 247:123–130, 2005.
153. D. Amikam and M. Y. Galperin. PilZ domain is part of the bacterial c-di-GMP binding protein. *Bioinformatics*, 22:3–6, 2006.
154. D. A. Ryjenkov, R. Simm, U. Römling, and M. Gomelsky. The PilZ domain is a receptor for the second messenger c-di-GMP. *J. Biol. Chem.*, 281:30310–30314, 2006.

155. M. Christen, B. Christen, M. G. Allan, M. Folcher, P. Jenö, S. Grzesiek, and U. Jenal. DgrA is a member of a new family of cyclic diguanosine monophosphate receptors and controls flagellar motor function in *Caulobacter crescentus*. *Proc. Natl. Acad. Sci. USA*, 104:4112–4117, 2007.
156. T. A. Ramelot, A. Yee, J. R. Court, A. Semesi, C. H. Arrowsmith, and M. A. Kennedy. NMR structure and binding studies confirm that PA4608 from *Pseudomonas aeruginosa* is a PilZ domain and a c-di-GMP binding protein. *Proteins: Struct. Funct. Bioinf.*, 66:266–271, 2007.
157. P. Aldridge, R. Paul, P. Goymer, P. Rainey, and U. Jenal. Role of the GGDEF regulator PleD in polar development of *Caulobacter crescentus*. *Mol. Microbiol.*, 47:1695–1708, 2003.
158. R. Paul, S. Abel, P. Wassmann, A. Beck, H. Heerklotz, and U. Jenal. Activation of the diguanylate cyclase PleD by phosphorylation-mediated dimerization. *J. Biol. Chem.*, 282:29170–29177, 2007.
159. A. Levi and U. Jenal. Holdfast formation in motile swarmer cells optimizes surface attachment during *Caulobacter crescentus* development. *J. Bacteriol.*, 188:5315–5318, 2006.
160. P. H. Viollier and L. Shapiro. Spatial complexity of mechanisms controlling a bacterial cell cycle. *Curr. Op. Microbiol.*, 7:572–578, 2004.
161. V. L. Robinson, D. R. Buckler, and A. M. Stock. A tale of two components: a novel kinase and a regulatory switch. *Nat. Struct. Biol.*, 7:626–633, 2000.
162. A. M. Stock, V. L. Robinson, and P. N. Goudreau. Two-component signal transduction. *Annu. Rev. Biochem.*, 69:183–215, 2000.
163. G. J. Burton, G. B. Hecht, and A. Newton. Roles of the histidine protein kinase PleC in *Caulobacter crescentus* motility and chemotaxis. *J. Bacteriol.*, 179:5849–5853, 1997.
164. R. J. Lewis, J. A. Brannigan, K. Muchová, I. Barák, and A. J. Wilkinson. Phosphorylated aspartate in the structure of a response regulator protein. *J. Mol. Biol.*, 294:9–15, 1999.

## Bibliography

---

165. D. Yan, H. S. Cho, C. A. Hastings, M. M. Igo, S.-Y. Lee, J. G. Pelton, V. Stewart, D. E. Wemmer, and S. Kustu. Beryll fluoride mimics phosphorylation of NtrC and other bacterial response regulators. *Proc. Natl. Acad. Sci. USA*, 96:14789–14794, 1999.
166. D. E. Wemmer and D. Kern. Beryll fluoride binding mimics phosphorylation of aspartate in response regulators. *J. Bacteriol.*, 187:8229–8230, 2005.
167. C. A. Hastings, S.-Y. Lee, H. S. Cho, D. Yan, S. Kustu, and D. E. Wemmer. High-resolution solution structure of the beryll fluoride-activated NtrC receiver domain. *Biochemistry*, 42:9081–9090, 2003.
168. S.-Y. Lee, H. S. Cho, J. G. Pelton, D. Yan, R. K. Henderson, D. S. King, L.-S. Huang, S. Kustu, E. A. Berry, and D. E. Wemmer. Crystal structure of an activated response regulator bound to its target. *Nat. Struct. Biol.*, 8:52–56, 2001.
169. C. M. Dyer and F. W. Dahlquist. Switched or not?: The structure of unphosphorylated CheY bound to the N terminus of FliM. *J. Bacteriol.*, 188:7354–7363, 2006.
170. C. Birck, L. Mourey, P. Gouet, B. Fabry, J. Schumacher, P. Rousseau, D. Kahn, and J.-P. Samama. Conformational changes induced by phosphorylation of the FixJ receiver domain. *Structure*, 7:1505–1515, 1999.
171. M. H. Knaggs, F. R. Salsbury Jr., M. H. Edgell, and J. S. Fetrow. Insights into correlated motions and long-range interactions in CheY derived from molecular dynamics simulations. *Biophys. J.*, 92:2062–2079, 2007.
172. J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten. Scalable molecular dynamics with NAMD. *J. Comput. Chem.*, 26:1781–1802, 2005.
173. J.-P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J. Chem. Phys.*, 23:327–341, 1977.
174. D. Frishman and P. Argos. Knowledge-based protein secondary structure assignment. *Proteins: structure, function and genetics*, 23:566–579, 1995.



175. W. Humphrey, A. Dalke, and K. Schulten. VMD - visual molecular dynamics. *J. Molec. Graphics*, 14:33–38, 1996.
176. A. Toro-Roman, T. Wu, and A. M. Stock. A common dimerization interface in bacterial response regulator KdpE and TorR. *Prot. Sci.*, 14:3077–3088, 2005.
177. A. Toro-Roman, T. R. Mack, and A. M. Stock. Structural analysis and solution studies of the activated regulatory domain of the response regulator ArcA: A symmetric dimer mediated by the  $\alpha 4$ - $\beta 5$ - $\alpha 5$  face. *J. Mol. Biol.*, 349:11–26, 2005.
178. P. Bachhawat, G. V. T. Swapna, G. T. Montelione, and A. M. Stock. Mechanism of activation for transcription factor PhoB suggested by different modes of dimerization in the inactive and active states. *Structure*, 13:1353–1363, 2005.
179. P. Bachhawat and A. M. Stock. Crystal structures of the receiver domain of the response regulator PhoP from *Escherichia coli* in the absence and presence of the phosphoryl analog beryll fluoride. *J. Bacteriol.*, 189:5987–5995, 2007.
180. C. Chennubhotla and I. Bahar. Signal propagation in proteins and relation to equilibrium fluctuations. *PLoS Comp. Biol.*, 3:e172, 2007.
181. J. A. Hardy and J. A. Wells. Searching for new allosteric sites in enzymes. *Curr. Op. Struct. Biol.*, 14:1–10, 2004.
182. N. Homeyer, A. H. C. Horn, H. Lanig, and H. Sticht. AMBER force-field parameters for phosphorylated amino acids in different protonation states: phosphoserine, phosphothreonine, phosphotyrosine, and phosphohistidine. *J. Mol. Model.*, 12:281–289, 2006.
183. E. S. Groban, A. Narayanan, and M. P. Jacobson. Conformational changes in protein loops and helices induced by post-translational phosphorylation. *PLoS Comp. Biol.*, 2:0238–0250, 2006.



# Appendix

## Curriculum vitae

### Personal Data

Address: Franziska F.-F. Schmid  
Gotthardstrasse 124  
4054 Basel

Email: Franziska.Schmid@unibas.ch

Nationality: Swiss

### Education

11/2004 – 11/2008 **PhD Studies, Department of Chemistry, University of Basel**  
- Thesis: Protein Allostery: Computational Characterization of Diguanylate Cyclase PleD  
- Supervisor: Prof. M. Meuwly, University of Basel  
- External Examiner: Prof. L. Nilsson, Karolinska Institutet, Sweden

10/2000 – 12/2008 **Courses in Pharmacy and Biology, University of Basel**

09/2000 – 10/2004 **Studies in Chemistry, University of Basel**  
- Thesis: Combining MD Simulations and NMR Data for Analysis of Hydrogen Bonding in Small Proteins and Refinement of Force Field Parameters  
- Supervisor: Prof. M. Meuwly

08/1996 – 05/2000 **Matura Typ. C (Math), Alte Kantonsschule Aarau**

### Publications

- **Direct Comparison of Experimental and Calculated NMR Scalar Coupling Constants for Force Field Validation and Adaptation**  
F. F.-F. Schmid, M. Meuwly.  
*J. Chem. Theory Comput.* **2008**, 4(11): 1949 – 1958
- **All-atom Simulations of Structures and Energetics of c-di-GMP-bound and free PleD**  
F. F.-F. Schmid, M. Meuwly.  
*J. Mol. Biol.* **2007**, 374(5):1270 – 1285
- **Correlation of Protein Structure and Dynamics to Scalar Couplings across Hydrogen Bonds**  
H.-J. Sass, F. F.-F. Schmid, S. Grzesiek.  
*J. Am. Chem. Soc.* **2007**, 129(18):5898 – 5903
- **Allosteric Control of Cyclic di-GMP Signaling**  
B. Christen, M. Christen, R. Paul, F. Schmid, M. Folcher, P. Jenoe, M. Meuwly, U. Jenal.  
*J. Biol. Chem.* **2006**, 281(42):32015 – 32024
- **Insulin: a model system for nanomedicine?**  
M. Koch, F. F.-F. Schmid, V. Zoete, M. Meuwly.  
*Future Medicine* **2006**, 1(3):373 – 377

### Poster & Oral Presentations at Conferences

- **Molecular Dynamics Simulations to Study the PleD Dimerization Process**  
Poster, 2008 Swiss Chemical Society Fall Meeting in Zurich, Switzerland
- **Investigation of Parameters Influencing  $^3J_{NC'}$ -couplings Across Hydrogen Bonds Using MD Simulations**  
Poster, 2007 Swiss Chemical Society Fall Meeting in Lausanne, Switzerland
- **Computational Characterization of the Feedback Mechanism in PleD**  
Presentation, 2006 Swiss Chemical Society Fall Meeting in Zurich, Switzerland
- **Computational Approach to Allosteric Mechanism: Feedback Inhibition in Guanylate Cyclase PleD**  
Poster, 2006 Gordon Research Conference “Computational Chemistry” in Les Diablerets, Switzerland
- **Computational Approach to Allosteric Mechanism: Feedback Inhibition in Guanylate Cyclase PleD**  
Poster, 2006 1<sup>st</sup> European Chemistry Congress in Budapest, Hungary
- **Functional Analysis of Diguanylate Cyclase PleD by Molecular Dynamics Simulations**  
Poster, 2005 15<sup>th</sup> IUPAB / 5<sup>th</sup> EBSA International Biophysics Congress in Montpellier, France

## Work Experience & Practical Training

2004 – 2008	<b>Computer Administration of the Department</b> , University of Basel - Setup and maintenance of Linux computer cluster - Technical support for students
2004 – 2008	<b>Supervision of Undergraduate Research Projects</b> , University of Basel
2007 – 2008	<b>Participating in “Women Into Industry (WIN)”</b> a joint mentoring program by the University Basel and Novartis
2006	<b>Participating in the 3ème Cycle Summer School “Computational Quantum Dynamics”</b>
2005	<b>Participating in a “Parallel Programming” Workshop</b> at the CSCS in Manno, Switzerland
2003	<b>Practical Training with Prof. B. Ernst</b> , Department of Pharmacy, University of Basel - Bioenzymatic and regular chemical synthesis of Sialyl Lexis <sup>X</sup> and analoga - Compound analysis using 2D-NMR spectra (COSY, HSQC)
1998 – 2002	<b>Private Tutoring in Math and French for Secondary School Students</b> (age 12 to 16)
2000	<b>Participating in SJF Research Week “Forschen in der Chemie”</b>
1999	<b>Participating in SJF Research Week “Elektrotechnik und Maschinenbau”</b>

## Personal Skills

Operating System	Proficient in use and administration of Linux and Windows
Software	Molecular dynamics packages (CHARMM, NAMD) Visualization programs (VMD, MolMol, Molden) Least-squares fitting program I-NoLLS <i>Ab-initio</i> package Gaussian
Programming	Experienced in Fortran77, Tcl/Tk, HTML, BASH Basic knowledge in MPI and Perl
Languages	German (native), English (fluent), French (conversational)